

# UC Irvine

## UC Irvine Previously Published Works

### Title

Multidimensional Schrödinger operators whose spectrum features a half-line and a Cantor set

### Permalink

<https://escholarship.org/uc/item/0tw3r218>

### Journal

Journal of Functional Analysis, 280(7)

### ISSN

0022-1236

### Authors

Damanik, D  
Fillman, J  
Gorodetski, A

### Publication Date

2021-04-01

### DOI

10.1016/j.jfa.2020.108911

Peer reviewed

# MULTIDIMENSIONAL SCHRÖDINGER OPERATORS WHOSE SPECTRUM FEATURES A HALF-LINE AND A CANTOR SET

DAVID DAMANIK, JAKE FILLMAN, AND ANTON GORODETSKI

**ABSTRACT.** We construct multidimensional Schrödinger operators with a spectrum that has no gaps at high energies and that is nowhere dense at low energies. This gives the first example for which this widely expected topological structure of the spectrum in the class of uniformly recurrent Schrödinger operators, namely the coexistence of a half-line and a Cantor-type structure, can be confirmed. Our construction uses Schrödinger operators with separable potentials that decompose into one-dimensional potentials generated by the Fibonacci sequence and relies on the study of such operators via the trace map and the Fricke-Vogt invariant. To show that the spectrum contains a half-line, we prove an abstract Bethe–Sommerfeld criterion for sums of Cantor sets which may be of independent interest.

## 1. INTRODUCTION

The spectral analysis of Schrödinger operators

$$H_V = -\Delta + V \quad \text{in } L^2(\mathbb{R}^d)$$

plays a central role in quantum mechanics. The most fundamental issue is the study of the spectrum of  $H_V$ ,

$$\sigma(H_V) = \{E : (H_V - E)^{-1} \text{ does not exist as a bounded operator}\}.$$

The simplest case,  $V \equiv 0$ , can be readily understood via the Fourier transform and one finds  $\sigma(H_0) = [0, \infty)$ . The classical Weyl theorem then shows that  $\sigma_{\text{ess}}(H_V) = [0, \infty)$  as long as  $V$  is a relatively compact perturbation of  $H_0 = -\Delta$ . A sufficient condition for that is, e.g.,  $V \in L^p(\mathbb{R}^d) + L^\infty_\varepsilon(\mathbb{R}^d)$  (i.e. for every  $\varepsilon > 0$ , we can write  $V = V_{1,\varepsilon} + V_{2,\varepsilon}$  with  $V_{1,\varepsilon} \in L^p(\mathbb{R}^d)$  and  $\|V_{2,\varepsilon}\|_\infty < \varepsilon$ ) with  $p \geq \max\{2, d/2\}$  if  $d \neq 4$  and  $p > 2$  if  $d = 4$ ; compare [48, Problem XIII.41]. In particular,  $\sigma(H_V)$  contains a half line if  $V$  is (locally sufficiently regular and) arbitrarily small near infinity.

On the other hand, the spectrum may have gaps at high energies when  $V$  is not arbitrarily small near infinity. For an explicit example, one may consider the case  $d = 1$  and  $V(x) = \cos x$ ; see [48, Example 1 in Section XIII.16]. More generally, a generic periodic potential  $V$  will give rise in one space dimension to a Schrödinger operator whose spectrum does not contain a half line [49]. However, this is a strictly one-dimensional phenomenon. If  $d \geq 2$  and  $V$  is periodic (i.e., the periodicity

---

*Date:* January 14, 2020.

D.D. was supported in part by NSF grant DMS–1700131 and by an Alexander von Humboldt Foundation research award.

A.G. was supported in part by Simons Fellowship (grant number 556910), NSF grant DMS–1855541, and by Laboratory of Dynamical Systems and Applications NRU HSE, grant of the Ministry of science and higher education of the RF ag. N 075-15-2019-1931.

vectors of  $V$  form a lattice), then  $\sigma(H_V)$  does contain a half line. This statement is known as the *Bethe-Sommerfeld conjecture*, which is in fact a theorem due to work of several authors; see [42, 50, 51, 52, 55] and references therein. In recent years, there has also been interest in analogs of the Bethe-Sommerfeld conjecture for discrete Schrödinger operators on periodic lattices [18, 20, 24, 30] and quantum graphs [19].

It is natural to take this further and attempt to establish the Bethe-Sommerfeld property (i.e., the presence of a half line in the spectrum) for more general potentials  $V$ . On the positive side, there has been recent work establishing this result for certain classes of almost periodic potentials in dimension two [26, 27]. However, there is also a negative result: it was shown in [9] that there are almost periodic potentials  $V$  in any dimension for which  $\sigma(H_V)$  is a generalized Cantor set, that is, a perfect set with empty interior. This result provides the first example of an almost periodic Schrödinger operator in dimension  $d \geq 2$  for which the spectrum does not contain a half line, and it provides the first example of a Schrödinger operator in dimension  $d \geq 2$  with any potential for which the spectrum is a generalized Cantor set.

One should nevertheless expect that for “typical” almost periodic Schrödinger operators in dimension  $d \geq 2$ , the spectrum contains a half line. On the other hand, the structure of the spectrum at low to medium energies is expected to be quite different. While there are absolutely no results in this direction, there is an extensive literature on almost periodic Schrödinger operators in one dimension and it is by now well understood that those operators tend to have Cantor spectrum; compare, for example, [1, 10, 17, 38]. It is not unreasonable to expect that the spectra of “typical” almost periodic Schrödinger operators in dimension  $d \geq 2$  do exhibit a Cantor structure at low and/or medium energies, in addition to the half line at high energies.

More generally one may ask whether there are any Schrödinger operators in dimension  $d \geq 2$  for which such a coexistence phenomenon, a Cantor structure and a half line in separate energy regimes, can be established. In fact, we were asked this precise question by Leonid Parnowski and the purpose of this paper is to provide an affirmative answer. Thus, our main result is the construction of the first examples of potentials in  $\mathbb{R}^d$  exhibiting such a coexistence phenomenon.

Our examples will not be almost periodic in the classical (Bohr/Bochner) sense, but they belong to a class of uniformly recurrent potentials, which is defined as follows:

**Definition 1.1.** *Let  $d \geq 1$ . A function  $V : \mathbb{R}^d \rightarrow \mathbb{R}$  is called uniformly recurrent if for every  $r > 0$ , there exists  $R > 0$  such that for every  $x_0 \in \mathbb{R}^d$  and every  $x_1 \in \mathbb{R}^d$ , there exists  $x_2 \in \mathbb{R}^d$  with  $0 < |x_1 - x_2| < R$  such that  $V(x - x_0) = V(x - x_2)$  for every  $x \in \mathbb{R}^d$  with  $|x| < r$ . The set of bounded uniformly recurrent  $V : \mathbb{R}^d \rightarrow \mathbb{R}$  is denoted by  $\mathcal{UR}_d$ .*

**Theorem 1.2.** *For every  $d \geq 2$ , there are potentials  $V \in \mathcal{UR}_d$  such that the spectrum of the associated Schrödinger operator in  $L^2(\mathbb{R}^d)$  has a Cantor structure at low energies and possesses no gaps at high energies. More precisely, there are  $E_0 < E_1$  such that  $\sigma(H_V) \cap (-\infty, E_0]$  is perfect, nonempty, and nowhere dense and  $\sigma(H_V) \cap [E_1, \infty) = [E_1, \infty)$ .*

**Remark 1.3.** (a) Uniform recurrence is related in spirit to classical almost periodicity. Both try to capture the fact that the sets of translates for which one observes suitable repetitions are relatively dense. The notions differ in the definition of a suitable repetition. Bohr almost periodicity requires global repetition up to a small  $L^\infty$  error, while uniform recurrence requires exact repetition in a neighborhood of a given size. In particular, neither notion implies the other.

(b) Uniform recurrence is a standard notion in discrete geometry, and especially in the study of aperiodically ordered point sets, which in turn are of interest in the mathematical study of quasicrystals; see [2] and references therein.

(c) The potentials  $V$  for which we prove these statements are completely explicit. They arise from a product construction based on the Fibonacci sequence, which may be expressed in terms of a discontinuous sampling function defined over an irrational rotation as follows:

$$\omega_n = \chi_{[1-\theta,1)}(n\theta \bmod 1), \quad \theta = \frac{\sqrt{5}-1}{2}.$$

(d) It would of course be of interest to exhibit almost periodic  $V : \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $d \geq 2$ , for which the spectrum of the associated Schrödinger operator has both a Cantor component and a half line. We have been unable to establish the existence of such potentials. The obstacles which appear to make this task difficult are discussed in Section 6; see Question 4 and the discussion following it.

As already indicated above, our level of understanding is vastly different in the cases  $d = 1$  and  $d \geq 2$ . Far more is known in the one-dimensional case, and in order to take advantage of that we will consider separable potentials  $V : \mathbb{R}^d \rightarrow \mathbb{R}$ , that is, potentials of the form

$$(1.1) \quad V(x_1, \dots, x_d) = V_1(x_1) + \dots + V_d(x_d)$$

with  $V_j : \mathbb{R} \rightarrow \mathbb{R}$ ,  $1 \leq j \leq d$ . Thus, in addition to the operator  $H_V$  in  $L^2(\mathbb{R}^d)$  of main interest, we may also consider the operators  $H_{V_j}$  in  $L^2(\mathbb{R})$ ,  $1 \leq j \leq d$ . It is well known that the spectra are related via the standard Minkowski sum of sets as follows:

$$(1.2) \quad \sigma(H_V) = \sigma(H_{V_1}) + \dots + \sigma(H_{V_d}) = \{E_1 + \dots + E_d : E_j \in \sigma(H_{V_j}), 1 \leq j \leq d\}.$$

See, for example, [47, Sections II.4 and VIII.10] and [47, Theorem II.10].

This suggests the strategy we will follow in the proof of Theorem 1.2. Choose one-dimensional potentials  $V_j$  such that the spectra  $\sigma(H_{V_j})$  are generalized Cantor sets. Then investigate the sum on the right-hand side of (1.2) and prove that there are  $E_0, E_1$  such that it is Cantor below  $E_0$  and a half line above  $E_1$ . It is well known that the study of sums of Cantor sets is difficult and can lead to a variety of outcomes. Such sums can indeed be Cantor sets or intervals, and certain mechanisms are known that will produce either outcome. The challenge will then be to choose the potentials  $V_j$  in such a way that the spectra  $\sigma(H_{V_j})$  have a structure that will allow us to apply both of these mechanisms simultaneously.

Once we have explained which features of Cantor sets we will need, we may explore these features in spectra of specific one-dimensional Schrödinger operators. Our one-dimensional potentials will be generated by the Fibonacci sequence. This is not a coincidence. This is currently (essentially) the only example for which the

desired features can be established. In particular, no bona fide almost periodic potential in one dimension is currently known for which similar results can be proved. Thus, in Subsection 3.1 we will discuss one-dimensional potentials generated by the Fibonacci substitution and recall some important concepts needed in the spectral analysis of the associated Schrödinger operators, namely the trace map, the Fricke-Vogt invariant, and the curve of initial conditions.

The remainder of the paper is organized as follows. In Section 2, we formulate and prove an abstract Bethe–Sommerfeld criterion for sums of extended Cantor sets. We single out this result, since it may be of independent interest outside our main application. In Section 3 we review some background about potentials generated by the Fibonacci substitution. In Section 4, we verify the conditions of the abstract Bethe–Sommerfeld criterion for the Fibonacci model and hence confirm the half-line portion of Theorem 1.2. In Section 5, we study the low-energy region, showing that the spectrum can be made to be a Cantor set in suitable parameter regions. Finally, in Section 6 we conclude with some questions raised by the present work that we regard as challenging and interesting.

**Remark 1.4.** (a) For notational simplicity, we give the proof of Theorem 1.2 in the case  $d = 2$ . The proof in the case  $d > 2$  is completely analogous. In fact, no changes are necessary in the choice of the underlying 1D potential to get the half-line portion of the spectrum, and only a minor change is necessary to get the Cantor portion of the spectrum (namely, one needs to replace the number  $1/2$  in (5.2) by  $1/d$ ).

(b) It is interesting to compare the structure of the spectrum of Schrödinger operators given by Theorem 1.2 with the structure of Lagrange and Markov spectra, the sets that appear naturally in the theory of Diophantine approximation; see [32] and references therein.

**Acknowledgment.** We are grateful to Leonid Parnowski for raising the question of whether examples with spectra exhibiting the topological structure discussed in this paper can be found and for his interest in this work.

## 2. ABSTRACT BETHE–SOMMERFELD CRITERION

Let us begin by recalling some terminology and results that will be useful. Let  $C \subset \mathbb{R}$  be a Cantor set (i.e.,  $C$  is bounded, perfect, and nowhere dense) and denote by  $I$  its convex hull. Any connected component of  $I \setminus C$  is called a *gap* of  $C$ . A *presentation* of  $C$  is given by an ordering  $\mathcal{U} = \{U_n\}_{n \geq 1}$  of the gaps of  $C$ . If  $u \in C$  is a boundary point of a gap  $U$  of  $C$ , we denote by  $K$  the connected component of  $I \setminus (U_1 \cup U_2 \cup \dots \cup U_n)$  (with  $n$  chosen so that  $U_n = U$ ) that contains  $u$  and write

$$\tau(C, \mathcal{U}, u) = \frac{|K|}{|U|}.$$

The *thickness*  $\tau(C)$  of  $C$  is given by

$$\tau(C) = \sup_{\mathcal{U}} \inf_u \tau(C, \mathcal{U}, u),$$

The following consequence of the Newhouse Gap Lemma [39, 40] is stated as [12, Lemma 6.2] and proved there:

**Lemma 2.1.** *Suppose  $C, K \subset \mathbb{R}$  are Cantor sets with  $\tau(C) \cdot \tau(K) > 1$ . Assume also that the size of the largest gap of  $C$  is not greater than the diameter of  $K$ , and the size of the largest gap of  $K$  is not greater than the diameter of  $C$ . Then,*

$$C + K = [\min C + \min K, \max C + \max K].$$

**Remark 2.2.** A particular consequence of Lemma 2.1 is the following: if  $C \subset \mathbb{R}$  is a Cantor set with  $\tau(C) > 1$ , then

$$C + C = [2 \min C, 2 \max C].$$

**Lemma 2.3** (Abstract Bethe-Sommerfeld Conditions). *Suppose  $K \subset \mathbb{R}$  is a closed set that has the following properties:*

- (1)  *$K$  contains Cantor sets  $K_1, K_2, \dots$  with disjoint convex hulls  $I_1, I_2, \dots$ , so that  $I_{n+1}$  lies to the right of  $I_n$  for every  $n \in \mathbb{N}$ .*
- (2) *For some  $\varepsilon > 0$ , we have  $\tau(K_n) > 1 + \varepsilon$  for all  $n \in \mathbb{N}$ .*
- (3) *For some  $A > a > 0$ , we have  $2A \geq |I_n| \geq A$  and  $\text{dist}(I_n, I_{n+1}) \leq a$  for all  $n \in \mathbb{N}$ .*

*Let  $\tilde{K} = f(K)$ , where  $f : \mathbb{R} \rightarrow \mathbb{R}$  is given by  $f(x) = x^2$ . Then  $\tilde{K} + \tilde{K}$  contains a half-line.*

*Proof.* Denote  $\tilde{K}_n = f(K_n)$ . Throughout the argument, we assume  $n$  is large enough that  $K_n \subseteq [0, \infty)$ . Let us show that for all sufficiently large  $n \in \mathbb{N}$ , we have  $\tau(\tilde{K}_n) \geq 1 + \frac{\varepsilon}{2}$ . Indeed, for any two intervals  $B = [u, v]$ ,  $U = [v, w]$  from  $I_n$  such that

$$\frac{|B|}{|U|} = \frac{v - u}{w - v} \geq 1 + \varepsilon,$$

we have

$$\begin{aligned} \frac{|f(B)|}{|f(U)|} &= \frac{v^2 - u^2}{w^2 - v^2} \\ &= \frac{v - u}{w - v} \cdot \frac{v + u}{w + v} \\ &> (1 + \varepsilon) \left( \frac{2v - 2A}{2v + 2A} \right) \\ &= (1 + \varepsilon) \left( 1 - \frac{2A}{v + A} \right) \\ &> 1 + \frac{\varepsilon}{2}, \end{aligned}$$

for  $v$  sufficiently large (i.e.  $n$  sufficiently large).

A similar estimate holds in the case of intervals  $B = [u, v]$ ,  $U = [w, u]$  (i.e., in the language of the Newhouse Gap Lemma, when the bridge is to the right of the gap). Therefore, for all sufficiently large  $n$ ,  $\tau(\tilde{K}_n) \geq 1 + \frac{\varepsilon}{2}$ , and thus the sum  $\tilde{K}_n + \tilde{K}_n$  is an interval by Remark 2.2.

Next, let us show that for all sufficiently large  $n$ , each of the Cantor sets  $\tilde{K}_n$  and  $\tilde{K}_{n+1}$  has diameter larger than the largest gap of the other. In that case Lemma 2.1 will imply that  $\tilde{K}_n + \tilde{K}_{n+1}$  is an interval. Indeed, suppose that  $I_n = [x, y]$  and  $I_{n+1} = [z, t]$ . By our assumptions, we have

$$(2.1) \quad A \leq y - x \leq 2A, \quad z - y \leq a, \quad A \leq t - z \leq 2A.$$

Since we already know that  $\tau(\tilde{K}_{n+1}) > 1$ , the largest gap of  $\tilde{K}_{n+1}$  is not greater than  $\frac{1}{3}(t^2 - z^2)$ , and the diameter of  $\tilde{K}_n$  is equal to  $y^2 - x^2$ . Using (2.1) three times, we have

$$\begin{aligned} y^2 - x^2 &= (y - x)(y + x) \\ &\geq A(x + y) \\ &\geq A(2y - 2A) \\ &\geq A(2(z - a) - 2A) \\ &= 2Az - (2Aa + 2A^2). \end{aligned}$$

Similarly, using (2.1) twice more yields

$$\begin{aligned} \frac{1}{3}(t^2 - z^2) &= \frac{1}{3}(t - z)(t + z) \\ &\leq \frac{1}{3}2A(t + z) \\ &\leq \frac{2}{3}A(2z + 2A) \\ &= \frac{4}{3}Az + \frac{4}{3}A^2. \end{aligned}$$

Therefore, for all sufficiently large values of  $n$ , the diameter of  $\tilde{K}_n$  is greater than the largest gap of  $\tilde{K}_{n+1}$ . Similarly one can show that for all sufficiently large values of  $n$ , the diameter of  $\tilde{K}_{n+1}$  is greater than the largest gap of  $\tilde{K}_n$ .

We therefore know that the sets  $J_n := \tilde{K}_n + \tilde{K}_n$  and  $J'_n := \tilde{K}_n + \tilde{K}_{n+1}$  are intervals for large  $n$ . Let us show that they cover a half line. It is enough to check that  $J_n$  is not disjoint from  $J'_n$ , and  $J'_n$  is not disjoint from  $J_{n+1}$ .

As before, denote  $I_n = [x, y]$ ,  $I_{n+1} = [z, t]$ . It follows from our discussion above that  $J_n = [2x^2, 2y^2]$ ,  $J_{n+1} = [2z^2, 2t^2]$ , and  $J'_n = [x^2 + z^2, y^2 + t^2]$ . To show that  $J_n$  is not disjoint from  $J'_n$  we need to check that  $2y^2 \geq x^2 + z^2$ . Using (2.1) twice in each term, we have

$$\begin{aligned} 2y^2 - (x^2 + z^2) &= (y - x)(y + x) + (y - z)(y + z) \\ &\geq A(y + x) - a(y + z) \\ &\geq A(2y - 2A) - a(2y + a) \\ &= 2y(A - a) - 2A^2 - a^2 \\ &\geq 0 \end{aligned}$$

if  $n$  is sufficiently large.

To show that  $J'_n$  is not disjoint from  $J_{n+1}$  we need to check that  $y^2 + t^2 \geq 2z^2$ . We have

$$\begin{aligned} y^2 + t^2 - 2z^2 &= (t^2 - z^2) + (y^2 - z^2) \\ &= (t - z)(t + z) + (y - z)(y + z). \end{aligned}$$

Using  $t - z = |I_{n+1}| \geq A$  and  $y \leq z \leq t$  to estimate the first term and  $z - y \leq a$  (twice) to estimate the second term, we arrive at

$$\begin{aligned} y^2 + t^2 - 2z^2 &\geq 2Ay - a(2y + a) \\ &\geq 2(A - a)y - a^2 \end{aligned}$$

$$\geq 0$$

if  $n$  is sufficiently large.

It follows that for  $n_0$  large enough, the set

$$\bigcup_{n \geq n_0} (J_n \cup J'_n),$$

which is contained in  $\tilde{K} + \tilde{K}$ , is a half line.  $\square$

### 3. BACKGROUND

**3.1. Trace Map and Fricke–Vogt Invariant.** Let us recall the dynamical setup that is canonically associated with any one-dimensional Schrödinger operator whose potential is generated by the Fibonacci substitution. We refer the reader to [6, 12] for more details and background.

The *trace map* is given by

$$T : \mathbb{R}^3 \rightarrow \mathbb{R}^3, \quad T(x, y, z) = (2xy - z, x, y).$$

Note that  $T$  is invertible with  $T^{-1}(x, y, z) = (y, z, 2yz - x)$ .

The *Fricke–Vogt invariant*

$$G(x, y, z) = x^2 + y^2 + z^2 - 2xyz - 1$$

obeys

$$(3.1) \quad G \circ T = G,$$

and hence  $T$  preserves the family of cubic surfaces

$$S_I = \{(x, y, z) \in \mathbb{R}^3 : G(x, y, z) = I\}.$$

The surface  $S_0$  is called the *Cayley cubic*.

It is natural to consider the restriction  $T_I$  of the trace map  $T$  to the invariant surface  $S_I$ . That is,  $T_I : S_I \rightarrow S_I$ ,  $T_I = T|_{S_I}$ . Denote by  $\Omega_I$  the set of points in  $S_I$  whose full two-sided orbit under  $T_I$  is bounded.

Let us recall that an invariant closed set  $\Lambda$  of a diffeomorphism  $f : M \rightarrow M$  is *hyperbolic* if there exists a splitting of the tangent space  $T_x M = E_x^u \oplus E_x^s$  at every point  $x \in \Lambda$  such that this splitting is invariant under  $Df$ , the differential  $Df$  exponentially contracts vectors from the stable subspaces  $\{E_x^s\}$ , and the differential of the inverse,  $Df^{-1}$ , exponentially contracts vectors from the unstable subspaces  $\{E_x^u\}$ . A hyperbolic set  $\Lambda$  of a diffeomorphism  $f : M \rightarrow M$  is *locally maximal* if there exists a neighborhood  $U$  of  $\Lambda$  such that

$$\Lambda = \bigcap_{n \in \mathbb{Z}} f^n(U).$$

It was shown in [4, 5, 11] that for every  $I > 0$ , the set  $\Omega_I$  is a locally maximal hyperbolic set of  $T_I : S_I \rightarrow S_I$ , which is homeomorphic to a Cantor set.

For the discussion of the high-energy regime later in this paper, special attention needs to be paid to small non-negative values of  $I$ , which may then be viewed as small perturbations of the case  $I = 0$ . For this reason, let us briefly discuss the latter special case.

Denote by  $\mathbb{S}$  the part of the Cayley cubic  $S_0$  inside the cube  $\{|x| \leq 1, |y| \leq 1, |z| \leq 1\}$ . The surface  $\mathbb{S}$  is homeomorphic to  $S^2$ , invariant, smooth everywhere except at the four points  $P_1 = (1, 1, 1)$ ,  $P_2 = (-1, -1, 1)$ ,  $P_3 = (1, -1, -1)$ , and



$P_4 = (-1, 1, -1)$ , where  $\mathbb{S}$  has conic singularities, and the trace map  $T$  restricted to  $\mathbb{S}$  is a factor of the hyperbolic automorphism of  $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$  given by

$$\mathcal{A}(\theta, \varphi) = (\theta + \varphi, \theta) \pmod{1}.$$

The semiconjugacy is given by the map

$$F : (\theta, \varphi) \mapsto (\cos 2\pi(\theta + \varphi), \cos 2\pi\theta, \cos 2\pi\varphi).$$

The map  $\mathcal{A}$  is hyperbolic, and is given by the matrix  $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ , which has eigenvalues

$$\mu = \frac{1 + \sqrt{5}}{2} \quad \text{and} \quad -\mu^{-1} = \frac{1 - \sqrt{5}}{2}.$$

The Markov partition for the map  $\mathcal{A} : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  shown in Figure 1 has already appeared in [5, 12]; for more details on Markov partitions for two-dimensional hyperbolic maps see [41, Appendix 2]). Its image under the map  $F : \mathbb{T}^2 \rightarrow \mathbb{S}$  is a Markov partition for the pseudo-Anosov map  $T_0 : \mathbb{S} \rightarrow \mathbb{S}$ .

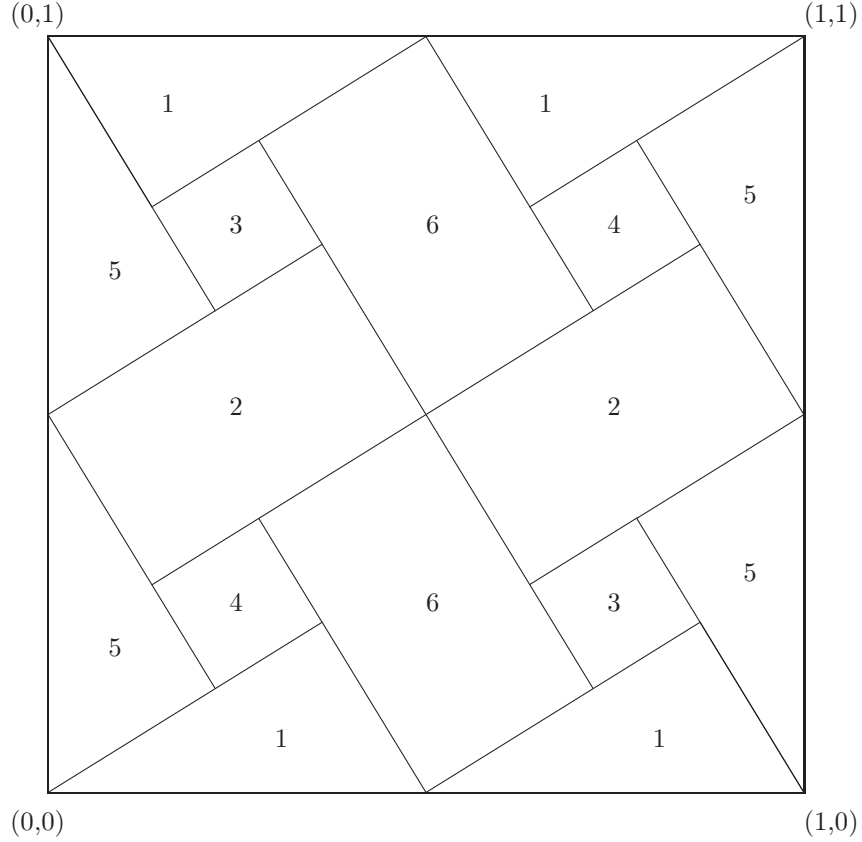


FIGURE 1. The Markov partition for the map  $\mathcal{A}$ .

So far we have discussed the dynamical setup which is tied to the trace map. Since all “second-order” one-dimensional operators (more precisely, operators for which there exists an  $\text{SL}(2, \mathbb{R})$  transfer matrix formalism) generated by the Fibonacci substitution lead to the same trace map, this dynamical setup is relevant to all of them. The specific choice of the operator in question, however, will affect the associated curve of initial conditions, which needs to be studied relative to the fixed dynamical setup. Let us discuss the operator and its associated curve of initial conditions relevant to this paper.

**3.2. Schrödinger Operators Associated with the Fibonacci Subshift.** Let us introduce the second-order operators that we will study. See also [8, 21]. First, we recall the notion of concatenation of real-valued functions defined on intervals. Given  $\ell_n \in \mathbb{R}_+ \stackrel{\text{def}}{=} (0, \infty)$  and  $f_n : [0, \ell_n) \rightarrow \mathbb{R}$  for each  $n \in \mathbb{Z}$ , define the *concatenation* of the sequence  $\{f_n\}_{n \in \mathbb{Z}}$  as follows. Put

$$(3.2) \quad t_n \stackrel{\text{def}}{=} \begin{cases} \sum_{j=0}^{n-1} \ell_j & n \geq 1 \\ 0 & n = 0 \\ -\sum_{j=n}^{-1} \ell_j & n \leq -1, \end{cases}$$

denote  $J_n = [t_n, t_{n+1})$ ,  $J = \bigcup_n J$  and define  $f : J \rightarrow \mathbb{R}$  by

$$(3.3) \quad f(x) = f_n(x - t_n), \quad \text{for each } x \in J_n.$$

In the present work, we will have  $\sum_{n < 0} \ell_n = \sum_{n > 0} \ell_n = \infty$  so that  $J = \mathbb{R}$ . Using a box to denote the position of the origin, we denote the concatenation by

$$(3.4) \quad f = \left( \cdots \mid f_{-2} \mid f_{-1} \mid \boxed{f_0} \mid f_1 \mid f_2 \mid \cdots \right).$$

Let  $\mathbf{A}$  be a finite set, called the *alphabet*. Equip  $\mathbf{A}$  with the discrete topology and endow  $\mathbf{A}^{\mathbb{Z}}$  with the corresponding product topology. The *left shift*

$$[T\omega](n) \stackrel{\text{def}}{=} \omega(n+1), \quad \omega \in \mathbf{A}^{\mathbb{Z}}, \quad n \in \mathbb{Z},$$

defines a homeomorphism from  $\mathbf{A}^{\mathbb{Z}}$  to itself. A subset  $\Omega \subseteq \mathbf{A}^{\mathbb{Z}}$  is called *T-invariant* if  $T^{-1}(\Omega) = \Omega$ . Any compact *T*-invariant subset of  $\mathbf{A}^{\mathbb{Z}}$  is called a *subshift*.

We can use the concatenation construction above to associate potentials (and hence Schrödinger operators) to elements of subshifts as follows. For each  $\alpha \in \mathbf{A}$ , we pick  $\ell_\alpha > 0$  and a real-valued function  $f_\alpha \in L^2[0, \ell_\alpha)$ . Then, for any  $\omega \in \mathbf{A}^{\mathbb{Z}}$ , we define the action of the continuum Schrödinger operator  $H_\omega$  in  $L^2(\mathbb{R})$  by

$$(3.5) \quad H_\omega = -\frac{d^2}{dx^2} + V_\omega,$$

where the potential  $V_\omega$  is given by

$$(3.6) \quad V_\omega = V_{\omega, \{f_\alpha\}} \stackrel{\text{def}}{=} \left( \cdots f_{\omega_{-2}} \mid f_{\omega_{-1}} \mid \boxed{f_{\omega_0}} \mid f_{\omega_1} \mid f_{\omega_2} \cdots \right).$$

These potentials belong to  $L^2_{\text{loc, unif}}(\mathbb{R})$  and hence each  $H_\omega$  defines a self-adjoint operator on a dense subspace  $L^2(\mathbb{R})$  in a canonical fashion.

In the present manuscript, the alphabet contains two symbols,  $\mathbf{A} \stackrel{\text{def}}{=} \{\mathbf{a}, \mathbf{b}\}$ . The Fibonacci substitution is the map

$$S(\mathbf{a}) = \mathbf{ab}, \quad S(\mathbf{b}) = \mathbf{a}.$$

This map extends by concatenation to  $A^*$ , the free monoid over  $A$  (i.e. the set of finite words over  $A$ ), as well as to  $A^{\mathbb{N}}$ , the collection of (one-sided) infinite words over  $A$ . There exists a unique element

$$u = \text{abaababa} \dots \in A^{\mathbb{N}}$$

with the property that  $u = S(u)$ . It is straightforward to verify that for  $n \in \mathbb{N}$ ,  $S^n(a)$  is a prefix of  $S^{n+1}(a)$ . Thus, one obtains  $u$  as the limit (in the product topology on  $A^{\mathbb{N}}$ ) of the sequence of finite words  $\{S^n(a)\}_{n \in \mathbb{N}}$ . With this setup, the Fibonacci subshift is defined to be the collection of two-sided infinite words with the same local factor structure as  $u$ , that is,

$$\Omega \stackrel{\text{def}}{=} \{\omega \in A^{\mathbb{Z}} : \text{every finite subword of } \omega \text{ is also a subword of } u\}.$$

Given  $\ell_a, \ell_b \in \mathbb{R}_+$  and real-valued functions  $f_a \in L^2[0, \ell_a)$ ,  $f_b \in L^2[0, \ell_b)$ , we consider the family of continuum Schrödinger operators  $\{H_\omega\}_{\omega \in \Omega}$  defined by (3.5) and (3.6). Since  $(\Omega, T)$  is a minimal dynamical system, one can verify that there is a uniform closed set  $\Sigma = \Sigma(f_a, f_b) \subset \mathbb{R}$  with the property that

$$(3.7) \quad \sigma(H_\omega) = \Sigma \text{ for every } \omega \in \Omega.$$

Of course, one can choose  $f_a$  and  $f_b$  in such a way that every  $V_\omega$  is a periodic potential (notice that as soon as  $V_{\omega_0}$  is periodic for a single  $\omega_0 \in \Omega$ , then every  $V_\omega$  is periodic by minimality). The main result of [8] is that this is the only possible obstruction to Cantor spectrum. Thus, we adopt the following assumption throughout the paper:

**Assumption 3.1.** The potential pieces  $f_a$  and  $f_b$  are chosen so that  $V_\omega$  is aperiodic for one  $\omega \in \Omega$  (hence for every  $\omega \in \Omega$  by minimality).

**Theorem 3.2** (Damanik–Fillman–Gorodetski [8]). *If  $f_a$  and  $f_b$  satisfy Assumption 3.1, then  $\Sigma(f_a, f_b)$  is a Cantor set of zero Lebesgue measure.*

**Remark 3.3.** In [8], the authors also assumed a condition on  $\{f_\alpha : \alpha \in A\}$  that they called *irreducibility*. This condition is defined so that the potentials satisfy the *simple finite decomposition property* (SFDP) from [29]. However, since our alphabet only has two letters, SFDP follows from aperiodicity and [29, Proposition 3.5].

In the sequel, it will be convenient to introduce an additional symbol: “ab”, and then to define  $\ell_{ab} = \ell_a + \ell_b$  and

$$f_{ab}(x) = \begin{cases} f_a(x) & 0 \leq x < \ell_a \\ f_b(x - \ell_a) & \ell_a \leq x < \ell_{ab}. \end{cases}$$

Given  $\alpha \in \tilde{A} \stackrel{\text{def}}{=} A \cup \{\text{ab}\}$ , consider the solutions of the differential equation  $-u''(x) + f_\alpha(x)u(x) = Eu(x)$  for  $E \in \mathbb{R}$ . Denote the solution obeying  $u(0) = 0$ ,  $u'(0) = 1$  (resp.,  $u(0) = 1$ ,  $u'(0) = 0$ ) by  $u_{\alpha,D}(\cdot, E)$  (resp.,  $u_{\alpha,N}(\cdot, E)$ ). Then, we set

$$M(\alpha, E) = \begin{pmatrix} u_{\alpha,N}(\ell_\alpha, E) & u_{\alpha,D}(\ell_\alpha, E) \\ u'_{\alpha,N}(\ell_\alpha, E) & u'_{\alpha,D}(\ell_\alpha, E) \end{pmatrix},$$

and

$$(3.8) \quad x_{-1}(E) = \frac{1}{2} \text{tr}(M(\mathbf{b}, E)),$$

$$(3.9) \quad x_0(E) = \frac{1}{2} \text{tr}(M(\mathbf{a}, E)),$$

$$(3.10) \quad x_1(E) = \frac{1}{2} \operatorname{tr} (M(\mathbf{ab}, E)).$$

It is straightforward to check that  $M(\mathbf{ab}, \cdot) = M(\mathbf{b}, \cdot)M(\mathbf{a}, \cdot)$ . The map  $\gamma : E \mapsto (x_1(E), x_0(E), x_{-1}(E))$  will be called the *curve of initial conditions*. Note that by (3.1), all the points  $T^n(x_1(E), x_0(E), x_{-1}(E))$  lie on the surface  $S_{I(E)}$ , where (with some abuse of notation) we set

$$I(E) = I(\gamma(E)).$$

It was shown in [8] that  $I(E) \geq 0$  for every  $E \in \Sigma$ .

For special choices of the local potential pieces, it is possible to compute the Fricke-Vogt invariant  $I(E)$  explicitly. For example, as in [8] let us consider the case  $\ell_a = \ell_b = 1$ ,  $f_a = \lambda \cdot \chi_{[0,1]}$  and  $f_b = 0 \cdot \chi_{[0,1]}$ , where  $\lambda > 0$  and let us recall the formulas obtained there. Clearly, the resulting potentials are aperiodic. One readily computes the traces as follows:

$$\begin{aligned} x_{-1}(E) &= \cos \sqrt{E}, \\ x_0(E) &= \cos \sqrt{E - \lambda}, \\ x_1(E) &= \cos \sqrt{E} \cos \sqrt{E - \lambda} - \frac{1}{2} \left( \sqrt{\frac{E}{E - \lambda}} + \sqrt{\frac{E - \lambda}{E}} \right) \sin \sqrt{E} \sin \sqrt{E - \lambda}, \end{aligned}$$

and therefore

$$(3.11) \quad I(E) = \frac{1}{4} \frac{\lambda^2}{E(E - \lambda)} \sin^2 \sqrt{E} \sin^2 \sqrt{E - \lambda}.$$

These expressions all define entire functions of  $E \in \mathbb{C}$ . For example, for  $E < 0$ , one has  $x_{-1}(E) = \cosh \sqrt{|E|}$ .

Let us now state our main 1D results which we will leverage to prove Theorem 1.2. Let  $\ell_a = \ell_b = 1$  and  $f_a = 0 \cdot \chi_{[0,1]}$  and  $f_b = \lambda \cdot \chi_{[0,1]}$  be as above, and write  $\Sigma_\lambda = \Sigma(f_a, f_b)$ .

**Theorem 3.4.** *For any  $\lambda \in \mathbb{R}$ ,  $\Sigma_\lambda + \Sigma_\lambda$  contains a half-line. More precisely, there exists  $E_1 = E_1(\lambda)$  such that  $\Sigma_\lambda + \Sigma_\lambda \supseteq [E_1, \infty)$ .*

**Theorem 3.5.** *There exists  $\lambda_0$  such that, for any  $\lambda \geq \lambda_0$ , there is  $E_0$  such that  $(-\infty, E_0] \cap (\Sigma_\lambda + \Sigma_\lambda)$  is a nonempty Cantor set of zero Lebesgue measure.*

*Proof of Theorem 1.2.* Note first that it is well known that each  $\omega \in \Omega$  is uniformly recurrent (where the definition of uniform recurrence for maps  $\omega : \mathbb{Z} \rightarrow \mathbb{R}$  is analogous to the one given in Definition 1.1 for maps  $V : \mathbb{R} \rightarrow \mathbb{R}$ ; namely, for every length  $m \in \mathbb{N}$ , there is a window size  $M \in \mathbb{N}$  such that each word of length  $M$  appearing in  $\omega$  contains every word of length  $m$  that appears in  $\omega$ ).

From this and the way the potentials  $V_\omega$  are constructed, it is then easily seen that each  $V_\omega$  is uniformly recurrent in the sense of Definition 1.1. The resulting 2-dimensional potential arising via (1.1) with  $V_1 = V_2 = V_\omega$  then inherits the uniform recurrence property from  $V_\omega$ . (Recall that we focus on the case  $d = 2$  in this proof and leave the details of the extension to the case  $d \geq 3$  to the reader.)

The statements about the spectrum are immediate from Theorem 3.4, Theorem 3.5, and Equation (1.2).  $\square$

## 4. THE HIGH-ENERGY REGIME

Our goal in this section is to prove Theorem 3.4, which states that in the high-energy regime, the spectrum of the separable 2D Schrödinger operators we consider contains no gaps. In this proof the existing (extensive) work on the Fibonacci Hamiltonian will be helpful, but not sufficient. We will actually have to delve into the technical details of the theory that has been developed and establish new results. For this reason, this section is somewhat demanding and while we attempt to provide a presentation that is as self-contained as possible, it may be necessary to consult some of the earlier papers (e.g., [8, 11, 12, 15, 22, 56]) for further background and motivation.

The curve of initial conditions for the degenerate case in which the potential vanishes identically (i.e., for the free Laplacian) contains the curve  $\Pi = \{(\cos(2t), \cos t, \cos t)\}$ ,  $t \geq 0$ . This curve corresponds to the line segment in Figure 1 that connects the point  $(0, 0)$  with the point  $(0.5, 0.5)$ . In particular, this curve is transversal to both the stable and unstable foliations of the trace map  $T_0$  on  $\mathbb{S}_0$ . The curve of initial conditions  $\Gamma(E)$  accumulates to the curve  $\Pi$  as  $E$  becomes large, if one sets  $t = \sqrt{E}$ . We would like to consider only the parts of  $\Gamma(E)$  that are bounded away from suitably small neighborhoods of the singularities, and to then show that on each connected piece of the curve, the set of points that have bounded positive semi-orbits (i.e., those that correspond to energies in the spectrum) form a Cantor set of large thickness. In order to formalize this construction we need to discuss the dynamics of the trace map  $T$  and to introduce some notation.

First consider the dynamics of  $T$  in a neighborhood of the singularity  $P_1 = (1, 1, 1)$ . Due to the symmetries of the trace map this will also provide information on the dynamics near the other singularities. Take  $r_0 > 0$  small and let  $O_{r_0}(P_1)$  be an  $r_0$ -neighborhood of the point  $P_1 = (1, 1, 1)$  in  $\mathbb{R}^3$ . Let us consider the set  $\text{Per}_2(T)$  of periodic points of  $T$  of period 2.

By a direct calculation one can show the following:

**Lemma 4.1.** *We have*

$$\text{Per}_2(T) = \left\{ (x, y, z) : x \in (-\infty, \tfrac{1}{2}) \cup (\tfrac{1}{2}, \infty), y = \frac{x}{2x-1}, z = x \right\}.$$

Notice that in a neighborhood of  $P_1$ , the intersection  $\mathfrak{J} \equiv \text{Per}_2(T) \cap O_{r_0}(P_1)$  is a smooth curve that is a normally hyperbolic with respect to  $T$  (see, e.g., Appendix 1 in [41] for the formal definition of normal hyperbolicity). Therefore, the local center-stable manifold  $W_{\text{loc}}^{\text{cs}}(I)$  and the local center-unstable manifold  $W_{\text{loc}}^{\text{cu}}(I)$  defined by

$$\begin{aligned} W_{\text{loc}}^{\text{cs}}(I) &= \{p \in O_{r_0}(P_1) : T^n(p) \in O_{r_0}(P_1) \text{ for all } n \in \mathbb{N}\}, \\ W_{\text{loc}}^{\text{cu}}(I) &= \{p \in O_{r_0}(P_1) : T^{-n}(p) \in O_{r_0}(P_1) \text{ for all } n \in \mathbb{N}\} \end{aligned}$$

are smooth two-dimensional surfaces. Also, the local strong stable manifold  $W_{\text{loc}}^{\text{ss}}(P_1)$  and the local strong unstable manifold  $W_{\text{loc}}^{\text{uu}}(P_1)$  of the fixed point  $P_1$ , defined by

$$\begin{aligned} W_{\text{loc}}^{\text{ss}}(P_1) &= \{p \in W_{\text{loc}}^{\text{cs}}(I) : T^n(p) \rightarrow P_1 \text{ as } n \rightarrow \infty\}, \\ W_{\text{loc}}^{\text{uu}}(P_1) &= \{p \in W_{\text{loc}}^{\text{cu}}(I) : T^{-n}(p) \rightarrow P_1 \text{ as } n \rightarrow \infty\}, \end{aligned}$$

are smooth curves.

The Markov partition for the pseudo-Anosov map  $T : \mathbb{S} \rightarrow \mathbb{S}$  can be extended to a Markov partition for the map  $T : S_I \rightarrow S_I$ . Namely, there are four singular points  $P_1 = (1, 1, 1)$ ,  $P_2 = (-1, -1, 1)$ ,  $P_3 = (1, -1, -1)$ , and  $P_4 = (-1, 1, -1)$  of  $\mathbb{S}$ .

The point  $P_1$  is a fixed point of  $T$ , and the points  $P_2, P_3, P_4$  form a periodic orbit of period 3. For small  $I > 0$ , on the surface  $S_I$  near  $P_1$  there is a hyperbolic orbit of the map  $T_I = T|_{S_I}$  of period 2, and near the orbit  $\{P_2, P_3, P_4\}$  there is a hyperbolic periodic orbit of period 6. Pieces of stable and unstable manifolds of these 8 periodic points form a Markov partition for  $T_I : S_I \rightarrow S_I$ . For  $I > 0$ , the elements of this Markov partition are disjoint. Let us denote these six rectangles (the elements of the Markov partition) by  $R_I^1, R_I^2, \dots, R_I^6$ . Let us also denote  $R_I = \bigcup_{i=1}^6 R_I^i$  and  $R = \bigcup_{0 \leq I} R_I$ . Finally, let us denote  $R^i = \bigcup_{0 \leq I} R_I^i$ ,  $i = 1, 2, \dots, 6$ .

Denote by  $\Omega^+$  the set of points in  $\{I \geq 0\} \subset \mathbb{R}^3$  with bounded positive semiorbits.

The set of fixed points of  $T^2$  in a neighborhood of  $P_1$  of size  $r_0 > 0$  is a smooth curve  $\text{Fix}(T^2, O_{r_0}(P_1)) = \text{Per}_2(T) \cap O_{r_0}(P_1)$ ; see Lemma 4.1 above. Each of the fixed points has one of the eigenvalues equal to 1, one greater than 1, and one smaller than 1 in absolute value. Therefore the curve  $\text{Fix}(T^2, O_{r_0}(P_1))$  is a normally hyperbolic manifold, and its stable set  $W^s(\text{Fix}(T^2, O_{r_0}(P_1)))$  is a smooth two dimensional surface; see [25]. The strong stable manifolds form a  $C^1$ -foliation of  $W^s(\text{Fix}(T^2, O_{r_0}(P_1)))$ ; see [46, Theorem B]. It is convenient to consider  $T^6$  since in this case each of the eight periodic points that were born from the singularities becomes a fixed point. Due to the symmetries of the trace map, the dynamics of  $T^6$  is the same in a neighborhood of each of the singularities  $P_1, P_2, P_3$ , and  $P_4$ .

In order to specify how to choose the pieces of the curve  $\Gamma(t)$ , let us consider an image  $T^{6m}(\Pi)$ . This is a curve that connects  $P_1$  with  $P_3$ , and each connected component of the intersection of this curve with one of the elements of the Markov partition is a curve that connects one stable boundary with another. Let us take the curve obtained from  $T^{6m}(\Pi)$  by cutting the first and the last pieces with respect to this splitting. If  $m$  is large, the preimage of this curve is  $\Pi$  with small pieces near the singularities removed; let us denote it by  $\tilde{\Pi}$ . The curve  $\tilde{\Pi}$  is parameterized by  $t \in [2\pi n + \alpha, (2n+1)\pi - \beta]$  (or by  $t \in [(2n+1)\pi + \beta, 2\pi n - \alpha]$ ) for some small  $\alpha, \beta > 0$ . In the curve  $\Gamma(t)$  one can find a sequence of pieces such that the gaps between them are small, and the  $T^{6m}$ -images of these pieces connect the points of the boundaries of the same  $R^i$  as  $T^{6m}(\tilde{\Pi})$ . Denote those pieces by  $\Gamma_n$ .

**Proposition 4.2.** *Let  $\{\Gamma_n\}$  be a sequence of curves in  $\mathbb{R}^3$  parameterized by  $t \in [2\pi n + \alpha_n, (2n+1)\pi - \beta_n] \equiv J_n$ ,  $\alpha_n \rightarrow \alpha, \beta_n \rightarrow \beta$ , such that*

- (1)  $\Gamma_n$  converges to  $\tilde{\Pi}$  in the  $C^2$  topology;
- (2)  $I(\Gamma_n(t)) > 0$  for all  $t \in J_n$ ;
- (3) for some uniform  $C > 0$ , we have  $\left| \frac{dI(\Gamma_n(t))}{dt} \right| \leq CI(\Gamma_n(t))$ ;
- (4) the images  $T^k(\Gamma_n)$  intersect exactly the same sets  $R^i$  (and in the same order) as  $T^k(\tilde{\Pi})$  for all  $k \geq 0$ .

*Then the set  $K_n = \{t \in J_n : \Gamma_n(t) \in \Omega^+\}$  is a Cantor set for all large values of  $n$ , and  $\tau(K_n) \rightarrow \infty$  as  $n \rightarrow \infty$ .*

The proof of Proposition 4.2 goes along similar lines as the proof of [12, Theorem 1.2], which asserts that the spectrum of the discrete Fibonacci Hamiltonian is a Cantor set of large thickness in the small coupling regime. The situation in Proposition 4.2 is more delicate since the curve of initial conditions (that is the meaning of  $\Gamma_n(t)$ ) does not belong to the same level surface  $S_I$  as in the discrete case. Nevertheless, the condition (3) from Proposition 4.2 ensures that the value of the Fricke-Vogt invariant  $I$  does not change too fast along the curve  $\Gamma_n(t)$ , and that

in turn allows one to modify the main steps of the proof of [12, Theorem 1.2] in order to derive Proposition 4.2. We carry out the details of the proof of Proposition 4.2 in Subsections 4.1, 4.2, 4.3, and 4.4.

**4.1. Choice of a Coordinate System in a Neighborhood of a Singular Point.** Due to the smoothness of the invariant manifolds of the curve of periodic points of period two described above, there exists a smooth change of coordinates  $\Phi : O_{r_0}(P_1) \rightarrow \mathbb{R}^3$  such that  $\Phi(P_1) = (0, 0, 0)$  and

- $\Phi(\mathcal{I})$  is a part of the line  $\{x = 0, z = 0\}$ ;
- $\Phi(W_{\text{loc}}^{\text{cs}}(\mathcal{I}))$  is a part of the plane  $\{z = 0\}$ ;
- $\Phi(W_{\text{loc}}^{\text{cu}}(\mathcal{I}))$  is a part of the plane  $\{x = 0\}$ ;
- $\Phi(W_{\text{loc}}^{\text{ss}}(P_1))$  is a part of the line  $\{y = 0, z = 0\}$ ;
- $\Phi(W_{\text{loc}}^{\text{uu}}(P_1))$  is a part of the line  $\{x = 0, y = 0\}$ .

Denote  $f = \Phi \circ T \circ \Phi^{-1}$ . Then,

$$A \equiv Df(0, 0, 0) = D(\Phi \circ T \circ \Phi^{-1})(0, 0, 0) = \begin{pmatrix} \xi^{-1} & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & \xi \end{pmatrix},$$

where  $\xi$  is the largest eigenvalue of the differential  $DT(P_1) : T_{P_1}\mathbb{R}^3 \rightarrow T_{P_1}\mathbb{R}^3$ ,

$$DT(P_1) = \begin{pmatrix} 2 & 2 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \xi = \frac{3 + \sqrt{5}}{2} = \mu^2.$$

Let us denote  $\mathfrak{S}_I = \Phi(S_I)$ . Then, away from  $(0, 0, 0)$ , the family  $\{\mathfrak{S}_I\}$  is a smooth family of surfaces,  $\mathfrak{S}_0$  is diffeomorphic to a cone, contains the lines  $\{y = 0, z = 0\}$  and  $\{x = 0, y = 0\}$ , and at each non-zero point on these lines, it has a quadratic tangency with a horizontal or vertical plane.

Due to the symmetries of the trace map, similar changes of coordinates exist in a neighborhood of each of the other singularities. Denote  $O_{r_0} = O_{r_0}(P_1) \cup O_{r_0}(P_2) \cup O_{r_0}(P_3) \cup O_{r_0}(P_4)$ .

The next statement shows that, roughly speaking, if a point stays in a neighborhood where normalizing coordinates are defined for  $N$  iterates, then it must be  $\xi^{-N}$ -close to the center-stable manifold of the curve of fixed points.

**Proposition 4.3** (Proposition 3.4 from [12]). *Given  $C_1 > 0, C_2 > 0, \xi > 1$ , there exist  $\delta_0 = \delta_0(C_1, C_2, \xi)$ ,  $N_0 = N_0(C_1, C_2, \xi, \delta_0) \in \mathbb{N}$ , and  $C^{**} > C^* > 0$  such that for any  $\delta \in (0, \delta_0)$ , the following holds.*

*Let  $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be a  $C^2$ -diffeomorphism such that*

- (i)  $\|f\|_{C^2} \leq C_1$ ;
- (ii) *the planes  $\{z = 0\}$  and  $\{x = 0\}$  are invariant under iterates of  $f$ ;*
- (iii) *every point of the line  $\{z = 0, x = 0\}$  is a fixed point of  $f$ ;*
- (iv) *at a point  $Q \in \{z = 0, x = 0\}$  we have*

$$Df(Q) = \begin{pmatrix} \xi^{-1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \xi \end{pmatrix};$$

- (v)  $\|Df(p) - A\| < \delta$  for every  $p \in \mathbb{R}^3$ , where

$$A = Df(Q) = \begin{pmatrix} \xi^{-1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \xi \end{pmatrix}.$$

Introduce the following cone fields in  $\mathbb{R}^3$ :

$$(4.1) \quad K_p^u = \{\mathbf{v} \in T_p\mathbb{R}^3, \mathbf{v} = \mathbf{v}_{xy} + \mathbf{v}_z : |\mathbf{v}_z| \geq C_2 \sqrt{|z_p|} |\mathbf{v}_{xy}|\},$$

$$(4.2) \quad K_p^{cu} = \{\mathbf{v} \in T_p\mathbb{R}^3, \mathbf{v} = \mathbf{v}_x + \mathbf{v}_{yz} : |\mathbf{v}_x| < 0.01 \xi^{-1} |\mathbf{v}_{yz}|\},$$

$$(4.3) \quad K_p^s = \{\mathbf{v} \in T_p\mathbb{R}^3, \mathbf{v} = \mathbf{v}_x + \mathbf{v}_{yz} : |\mathbf{v}_x| \geq C_2 \sqrt{|x_p|} |\mathbf{v}_{yz}|\},$$

$$(4.4) \quad K_p^{cs} = \{\mathbf{v} \in T_p\mathbb{R}^3, \mathbf{v} = \mathbf{v}_z + \mathbf{v}_{xy} : |\mathbf{v}_z| < 0.01 \xi^{-1} |\mathbf{v}_{xy}|\}.$$

Suppose that for a finite orbit  $p_0, p_1, p_3, \dots, p_N$ , we have

$$(p_0)_x \geq 1, (p_1)_x < 1, (p_N)_z \geq 1, (p_{N-1})_z < 1,$$

and there are curves  $\gamma_0$  and  $\gamma_N$  such that  $\gamma_0$  connects  $p_0$  with  $W^{ss}(Q)$  and is tangent to both cone fields  $K^u$  and  $K^{cu}$ , and  $\gamma_N$  connects  $p_N$  with  $W^{uu}(Q)$  and is tangent to both cone fields  $K^s$  and  $K^{cs}$ .

Then

$$\begin{aligned} C^* \xi^{-N} &\leq |(p_0)_z| \leq C^{**} \xi^{-N}, \quad \text{and} \\ C^* \xi^{-N} &\leq |(p_N)_x| \leq C^{**} \xi^{-N}. \end{aligned}$$

From the proof of Proposition 4.3 (i.e., of [12, Proposition 3.4]) one can extract the following statement:

**Lemma 4.4.** *In the setting of Proposition 4.3, denote  $b_k = \text{dist}(p_k, Q)$ . There exists a constant  $C' > 0$  (independent of  $k$  or  $N$ ) such that the following holds:*

*If  $k < N/2$ , then  $b_k \leq C'(\xi - \delta)^{-k}$ .*

*If  $k \geq N/2$ , then  $b_k \leq C'(\xi - \delta)^{-N+k}$ .*

Other statements from [12] we will need are the following:

**Proposition 4.5** (Proposition 3.15 from [12]). *Given  $C_1 > 0, C_2 > 0, \lambda > 1$ , there exist  $\delta_0 = \delta_0(C_1, C_2, \lambda)$ ,  $N_0 \in \mathbb{N}, N_0 = N_0(C_1, C_2, \lambda, \delta_0) \in \mathbb{N}$ , and  $\tilde{C} > 0$  such that for any  $\delta \in (0, \delta_0)$ , the following holds.*

*Under the conditions of and with the notation from Proposition 4.3, suppose that  $\mathbf{v} \in T_{p_0}\mathbb{R}^3, \mathbf{v} \in K_{p_0}^u$ . Then  $|Df_{p_0}^N(\mathbf{v})| \geq \tilde{C} \xi^{N/2} |\mathbf{v}|$ .*

Let us denote  $\mathbf{v}_k = Df^k(\mathbf{v}), k = 0, 1, \dots, N$ , and  $D_k = |(\mathbf{v}_k)_z|, d_k = |(\mathbf{v}_k)_{xy}|$ .

**Lemma 4.6** (Lemma 3.16 from [12]). *There exists  $k^*$  such that  $d_k \geq D_k$  for all  $k \leq k^*$ , and  $d_k < D_k$  for all  $k > k^*$ .*

**Lemma 4.7** (Lemma 3.17 from [12]). *There is a constant  $C_9$  independent of  $N$  such that*

$$\xi^{k^*} \leq C_9 \xi^{N/2}.$$

**Lemma 4.8.** *There exists  $C_6 > 0$  such that for all  $k \leq k^*$  we have*

$$C_6^{-1} \leq |(\mathbf{v}_k)_y| \leq C_6.$$

*Proof of Lemma 4.8.* From the proof of Lemma 4.6 (i.e., [12, Lemma 3.16]) one can immediately extract that  $|(\mathbf{v}_k)_y| \leq d_k \leq C_6$  for some constant  $C_6 > 0$  for all  $k \leq k^*$ . At the same time, since  $\|f\|_{C^2} \leq C_1$ , and

$$A = Df(Q) = \begin{pmatrix} \xi^{-1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \xi \end{pmatrix},$$



we have

$$|(\mathbf{v}_{k+1})_y| \geq |(\mathbf{v}_k)_y| - \min(\delta, Cb_k)\|\mathbf{v}_k\| \geq (1 - C^*b_k)|(\mathbf{v}_k)_y|,$$

where  $C^*$  is a constant independent of  $k < k^*$  and  $N$ . Hence, taking into account Lemma 4.4, we have

$$|(\mathbf{v}_k)_y| \geq \left[ \prod_{i=1}^k (1 - C^*b_i) \right] |(\mathbf{v}_0)_y| \geq C_6^{-1}$$

for some constant  $C_6 > 0$ .  $\square$

**Proposition 4.9** (Proposition 3.18 from [12]). *Given  $C_1 > 0, C_2 > 0, C_3 > 0, \xi > 1$ , there exist  $\delta_0 = \delta_0(C_1, C_2, C_3, \xi)$ ,  $N_0 = N_0(C_1, C_2, C_3, \lambda, \delta_0) \in \mathbb{N}$ , and  $C > 0$  such that for any  $\delta \in (0, \delta_0)$  and any  $\Delta > 0$ , the following holds.*

*Under the conditions of and with the notation from Proposition 4.3, suppose that the curve  $\gamma_0$  has a curvature bounded by  $C_3$ . Suppose also that for the points  $p = (x_p, y_p, z_p)$  and  $q = (x_q, y_q, z_q)$ , the following holds:*

- (1)  $p, q \in \gamma_0$ ;
- (2) For some  $N \geq N_0$  both  $f^N(p)$  and  $f^N(q)$  have  $z$ -coordinates larger than 1, and both  $f^{N-1}(p)$  and  $f^{N-1}(q)$  have  $z$ -coordinates not greater than 1;
- (3)  $\text{dist}(f^N(p), f^N(q)) = \Delta$ .

Denote  $p_k = f^k(p)$ ,  $q_k = f^k(q)$ ,  $k = 0, \dots, N$ . Let  $\mathbf{v} \in T_p\mathbb{R}^3$  and  $\mathbf{w} \in T_q\mathbb{R}^3$  be vectors tangent to  $\gamma_0$  and denote  $\mathbf{v}_k = Df^k(v)$  and  $\mathbf{w}_k = Df^k(w)$ ,  $k = 0, \dots, N$ . Let  $\alpha_k$  be the angle between  $\mathbf{v}_k$  and  $\mathbf{w}_k$ .

Then,

$$(4.5) \quad \sum_{k=0}^N \alpha_k < C\Delta \quad \text{and} \quad \sum_{k=0}^N \text{dist}(p_k, q_k) < C\Delta.$$

We will need to modify Proposition 4.9.

**Definition 4.10.** For any points  $p, q$  and any nonzero vectors  $\mathbf{v}_p, \mathbf{v}_q$ , define

$$(4.6) \quad \mathfrak{F}(p, q, \mathbf{v}_p, \mathbf{v}_q) \equiv \frac{\angle(\mathbf{v}_p, \mathbf{v}_q)}{\text{dist}(p, q)}.$$

**Definition 4.11.** For any  $C^2$  smooth curve  $\gamma$ , define

$$(4.7) \quad \mathfrak{F}(\gamma) = \max \mathfrak{F}(p, q, \mathbf{v}_p, \mathbf{v}_q),$$

where  $\mathbf{v}_p$  and  $\mathbf{v}_q$  are unit vectors tangent to  $\gamma$  at the points  $p$  and  $q$ , and the maximum is taken over all pairs of distinct points  $p, q \in \gamma$ .

**Remark 4.12.** Notice that  $\mathfrak{F}(\gamma)$  is the maximum of the curvature of the curve  $\gamma$  over all its points.

Here is the formal statement that we need:

**Proposition 4.13** (Modified version of Proposition 4.9). *Given  $C_1 > 0, C_2 > 0, \xi > 1$ , there exist  $\delta_0 = \delta_0(C_1, C_2, \xi)$ ,  $N_0 = N_0(C_1, C_2, \xi, \delta_0) \in \mathbb{N}$ , and  $C > 0$  such that for any  $\delta \in (0, \delta_0)$  and any  $\Delta > 0$ , the following holds.*

*Under the conditions of and with the notation from Proposition 4.3, suppose that for the points  $p = (x_p, y_p, z_p)$  and  $q = (x_q, y_q, z_q)$ , the following holds:*

- (1)  $p, q \in \gamma_0$ ;

- (2) for some  $N \geq N_0$ , both  $f^N(p)$  and  $f^N(q)$  have  $z$ -coordinates larger than 1, and both  $f^{N-1}(p)$  and  $f^{N-1}(q)$  have  $z$ -coordinates not greater than 1;  
 (3)  $\text{dist}(f^N(p), f^N(q)) = \Delta$ .

Denote  $p_k = f^k(p)$ ,  $q_k = f^k(q)$ ,  $k = 0, \dots, N$ . Let  $\mathbf{v} \in T_p \mathbb{R}^3$  and  $\mathbf{w} \in T_q \mathbb{R}^3$  be vectors tangent to  $\gamma_0$  and denote  $\mathbf{v}_k = Df^k(\mathbf{v})$  and  $\mathbf{w}_k = Df^k(\mathbf{w})$ ,  $k = 0, \dots, N$ . Let  $\alpha_k$  be the angle between  $\mathbf{v}_k$  and  $\mathbf{w}_k$ .

Denote  $\gamma_k = f^k(\gamma_0)$ , and set  $\mathfrak{F}_k = \mathfrak{F}(\gamma_k)$ ,  $k = 0, \dots, N$ .

Then,

$$(4.8) \quad \sum_{k=0}^N \text{dist}(p_k, q_k) < C\Delta, \quad \text{and}$$

$$(4.9) \quad \mathfrak{F}_N < \xi^{-N/4} \mathfrak{F}_0 + C.$$

Moreover, for any  $C_3 > 0$  there exists  $\hat{C} > 0$  such that if  $\mathfrak{F}_0 < C_3$ , then

$$(4.10) \quad \sum_{k=0}^N \alpha_k < \hat{C}\Delta.$$

*Proof of Proposition 4.13.* The inequalities (4.8) and (4.10) follow from Proposition 4.9, one just needs to observe that in the proof of Proposition 4.9 (i.e., of [12, Proposition 3.18]) the estimates on  $\sum_{k=0}^N \text{dist}(p_k, q_k)$  do not use any bound on the curvature of  $\gamma_0$ . Hence we only need to prove (4.9).

Notice that it is enough to prove Proposition 4.13 in the case when the points  $p$  and  $q$  are arbitrarily close to each other. This follows from the fact that in Definition 4.11 one can take the maximum over those pairs of points that are sufficiently close, compare with Remark 4.12.

Denote by  $\Gamma$  the piece of the curve  $\gamma_0$  between the points  $p_0$  and  $q_0$ , and set  $\Gamma_k = f^k(\Gamma)$ ,  $k = 0, 1, 2, \dots, N$ . Denote  $\mu_k = |\Gamma_k|$ . Due to the remark above we can assume that  $\mu_0 \ll 1$ , and that for any vector tangent to  $\Gamma$ , the value of  $k^*$  is the same.

We will need to use the following elementary estimates, whose proofs we leave to the reader.

**Lemma 4.14.** *Given  $\alpha_0 > 0$ ,  $\tau > 1$ ,  $\delta > 0$ , suppose that*

$$\alpha_{k+1} \leq \tau \alpha_k + \delta, \quad k = 0, 1, \dots$$

*Then  $\alpha_k \leq \tau^k \left( \alpha_0 + \frac{\delta}{\tau-1} \right)$ .*

**Lemma 4.15.** *Given  $\alpha_0 > 0$ ,  $0 < t < 1$ ,  $\Delta > 0$ , suppose that*

$$\alpha_{k+1} \leq t \alpha_k + \Delta, \quad k = 0, 1, \dots$$

*Then  $\alpha_k \leq t^k \alpha_0 + \frac{\Delta}{1-t}$ .*

Recall that  $\mathbf{v}_k$  and  $\mathbf{w}_k$  are the vectors tangent to  $\Gamma_k$  at  $p_k = f^k(p_0)$  and  $q_k = f^k(q_0)$ , and  $\alpha_k = \angle(\mathbf{v}_k, \mathbf{w}_k)$ .

We know that for  $k \leq k^*$ , one has

$$\alpha_{k+1} \leq (\xi + \delta) \alpha_k + C_1 \mu_k \leq (\xi + \delta) \alpha_k + C_1 C_6 \mu_0,$$

since Lemma 4.8 is applicable in this case.

We also know that for some  $m \in \mathbb{N}$  independent of  $N$ , and for all  $k > k^* + m$ , we have

$$\alpha_{k+1} \leq (\xi^{-1} + \delta)\alpha_k + C_1\mu_k \leq (\xi^{-1} + \delta)\alpha_k + C_1\mu_N, \quad \text{and} \quad \mu_{k+1} \geq (\xi - 2\delta)\mu_k.$$

Therefore, Lemma 4.14 gives

$$\alpha_{k^*} \leq (\xi + \delta)^{k^*} \left( \alpha_0 + \frac{C_1 C_6 \mu_0}{\xi + \delta - 1} \right).$$

We also have

$$\alpha_{k^*+m} \leq C_1^m \alpha_{k^*} + m C_1^m \mu_0.$$

Therefore, due to Lemma 4.15 we have

$$\alpha_N \leq (\xi^{-1} + 2\delta)^{N/2} (C_1^m \alpha_{k^*} + m C_1^m \mu_0) + \frac{\mu_N}{1 - \xi^{-1} - 2\delta}.$$

Combining these inequalities, we get

$$\begin{aligned} \mathfrak{F}(p_N, q_N, \mathbf{v}_N, \mathbf{w}_N) &= \frac{\alpha_N}{\mu_N} \leq \\ &\text{const} \cdot \frac{(\xi^{-1} + 2\delta)^{N/2} (\xi + \delta)^{k^*}}{(\xi - 2\delta)^{N/2}} \mathfrak{F}(p_0, q_0, \mathbf{v}_0, \mathbf{w}_0) + \text{const} \cdot \frac{(\xi^{-1} + 2\delta)^{N/2}}{(\xi - 2\delta)^{N/2}} + \text{const}, \end{aligned}$$

and (4.9) follows.  $\square$

**4.2. Distortion Property: Estimate of the Gap Sizes.** Let us consider  $\Gamma_n$ . Its image  $T^{6m}(\Gamma_n)$  intersects  $R$  in a finite number of closed curves. The gaps between these curves correspond to the gaps in the parameter space  $J_n$  that we will call the gaps of order one. Denote

$$M_n^- = \min_{t \in J_n} I(\Gamma_n(t)), \quad M_n^+ = \max_{t \in J_n} I(\Gamma_n(t)).$$

The assumption (1) from Proposition 4.2 implies that  $M_n^-, M_n^+ \rightarrow 0$  as  $n \rightarrow \infty$ , and assumption (3) implies that for some uniform constant  $C_1 > 1$  that is independent of  $n$ , the size of any gap of order one is between  $C_1^{-1} \sqrt{M_n^-}$  and  $C_1 \sqrt{M_n^+}$ .

Those gaps in  $J_n$  that are formed by the intersection of  $T^{6m+(k-1)}(\Gamma_n)$  with the complement of  $R$  and do not have order less than  $k$  will be called gaps of order  $k$ . It is clear that every gap in  $\Gamma_n \cap \Omega^+$  has some finite order. Therefore we have ordered all the gaps.

Consider some gap  $\mathfrak{l}_G \subset J_n$  of order  $k$ . A *bridge* that corresponds to this gap is a connected component of the complement of the union of all gaps of order  $\leq k$  next to the gap. There are two bridges that correspond to the chosen gap, take one of them, and denote it by  $\mathfrak{l}_B$ . Now let us consider  $\mathfrak{L}_G \equiv \Gamma_n^{-1} \circ T^{6m+(k-1)}(\Gamma_n(\mathfrak{l}_G))$  and  $\mathfrak{L}_B \equiv \Gamma_n^{-1} \circ T^{6m+(k-1)}(\Gamma_n(\mathfrak{l}_B))$ . By the definition of the order  $k$  of the gap we know that

$$(4.11) \quad C_3 \sqrt{M_n^-} \leq \frac{|\mathfrak{L}_G|}{|\mathfrak{L}_B|} \leq C_4 \sqrt{M_n^+}$$

for suitable constants  $C_3$  and  $C_4$  independent of  $n$ .

**Proposition 4.16** (Analog of Proposition 3.11 from [12]). *There is a constant  $K > 1$  independent of the choice of the gap and of  $n$  such that*

$$K^{-1} \frac{|\mathfrak{l}_G|}{|\mathfrak{l}_B|} \leq \frac{|\mathfrak{L}_G|}{|\mathfrak{L}_B|} \leq K \frac{|\mathfrak{l}_G|}{|\mathfrak{l}_B|}.$$

After some preparatory work in Subsection 4.3, we will prove Proposition 4.16 in Subsection 4.4. In the meantime, let us observe that the distortion property elucidated in Proposition 4.16 suffices to establish Proposition 4.2.

*Proof of Proposition 4.2.* The statement follows from Proposition 4.16, inequality (4.11), and the fact that  $M_n^-, M_n^+ \rightarrow 0$  as  $n \rightarrow \infty$ .  $\square$

**4.3. Preliminary Estimates.** Let us denote by  $U_\nu$  the bounded connected component of  $\left(\bigcup_{0 \leq I \leq \nu} S_I\right) \setminus O_{r_2}$ . Let us denote by  $\pi_I$  the orthogonal projection of  $S_I \cap U_\nu$  to  $\mathbb{S} \setminus O_{r_2}$ . Notice that  $F^{-1}(\mathbb{S} \setminus O_{r_2})$  is a torus without small neighborhoods of the preimages of the singularities. We can define

$$F_{[0, \nu]} : U_\nu \rightarrow \mathbb{T}^2 \times [0, \nu]$$

by

$$F_{[0, \nu]}(x) = (F^{-1}(\pi_{I(x)}(x)), I(x)),$$

and

$$\tilde{T} : F_{[0, \nu]}(U_\nu) \rightarrow \mathbb{T}^2 \times [0, \nu]$$

by

$$\tilde{T}(\tilde{x}, I) = (F^{-1} \circ \pi_I \circ T \circ \pi_I^{-1} \circ F(\tilde{x}), I).$$

In this case  $\tilde{T}$  is  $C^2$ -close to  $\mathcal{A} \times id : \mathbb{T}^2 \times [0, \mu] \rightarrow \mathbb{T}^2 \times [0, \mu]$  if  $\nu$  is small, and

$$\tilde{T} = H^{-1} \circ T \circ H,$$

where  $H(x, I) = \pi_{I(x)}^{-1} \circ F(x)$ , that is,  $\tilde{T}$  and  $T$  are smoothly conjugate. Set

$$(4.12) \quad \tilde{C} := \max(\|H\|_{C^2}, \|H^{-1}\|_{C^2}, \|\Phi\|_{C^2}, \|\Phi^{-1}\|_{C^2}, \|H \circ \Phi\|_{C^2}, \|(H \circ \Phi)^{-1}\|_{C^2}).$$

Informally speaking, the constant  $\tilde{C}$  gives an upper bound on the distortion induced by any change of coordinates that we may want to consider.

Define a cone field in  $\mathbb{T}^2 \times [0, \nu]$  using stable-unstable directions of the map  $\mathcal{A}$  as  $x$  and  $y$  coordinates:

$$K_{(x, y, I)}^{\text{cu}} = \{\bar{v} = (v_x, v_y, v_I) \mid |v_y| > 100|v_x|, |v_I| < C_0 I\}.$$

Due to assumption (3) in Proposition 4.2 and the fact that the curve  $\Pi$  is transversal to the stable foliation on  $\mathbb{S}$ , one can choose  $C_0 > 0$  and  $k_0 \in \mathbb{N}$  in such a way that  $\tilde{T}^k \circ H^{-1}(\Gamma_n)$  is tangent to  $K^{\text{cu}}$  for all  $k \geq k_0$ .

**Lemma 4.17** (Analog of Lemma 3.20 from [12]). *For  $\nu > 0$  small enough, there exists  $\eta \in (0, 1)$  such that for any  $p, q \in F_{[0, \nu]}(U_\nu)$  and unit vectors  $\mathbf{v}_p \in K_p^{\text{cu}}$ ,  $\mathbf{v}_q \in K_q^{\text{cu}}$ , we have*

$$\angle(D\tilde{T}_p(\mathbf{v}_p), D\tilde{T}_q(\mathbf{v}_q)) \leq \eta \angle(\mathbf{v}_p, \mathbf{v}_q) + 2\|\tilde{T}\|_{C^2} \text{dist}(p, q).$$

*Proof of Lemma 4.17.* If  $\nu$  is small, then  $\tilde{T}$  is  $C^2$ -close to the map  $\mathcal{A} \times id$ . In particular, for any point  $p \in F_{[0, \nu]}(U_\nu)$  and any vectors  $\mathbf{v}_1, \mathbf{v}_2 \in K_p^{\text{cu}}$ ,

$$\angle(D\tilde{T}_p(\mathbf{v}_1), D\tilde{T}_p(\mathbf{v}_2)) \leq \eta \angle(\mathbf{v}_1, \mathbf{v}_2),$$

where  $\eta \in (0, 1)$  can be chosen uniformly for all  $p \in F_{[0, \nu]}(U_\nu)$ . Therefore we have

$$\begin{aligned} \angle(D\tilde{T}_p(\mathbf{v}_p), D\tilde{T}_q(\mathbf{v}_q)) &\leq \angle(D\tilde{T}_p(\mathbf{v}_p), D\tilde{T}_p(\mathbf{v}_q)) + \angle(D\tilde{T}_p(\mathbf{v}_q), D\tilde{T}_q(\mathbf{v}_q)) \\ &\leq \eta \angle(\mathbf{v}_p, \mathbf{v}_q) + 2\|D\tilde{T}_p(\mathbf{v}_q) - D\tilde{T}_q(\mathbf{v}_q)\| \\ &\leq \eta \angle(\mathbf{v}_p, \mathbf{v}_q) + 2\|\tilde{T}\|_{C^2} \text{dist}(p, q), \end{aligned}$$

as claimed.  $\square$

**Lemma 4.18** (Analog of Lemma 3.22 from [12]). *For  $p, q \in F_{[0, \nu]}(U_\nu)$ ,  $p \neq q$ , and vectors  $\mathbf{v}_p \in K_p^{\text{cu}}$ ,  $\mathbf{v}_q \in K_q^{\text{cu}}$ , consider the function  $\mathfrak{F}(p, q, \mathbf{v}_p, \mathbf{v}_q)$ . Suppose that  $p$  and  $q$  belong to a curve that is tangent to the cone field  $K^{\text{cu}}$ . Then*

$$\mathfrak{F}(\tilde{T}(p), \tilde{T}(q), D\tilde{T}_p(\mathbf{v}_p), D\tilde{T}_q(\mathbf{v}_q)) \leq \eta \mathfrak{F}(p, q, \mathbf{v}_p, \mathbf{v}_q) + 2\|\tilde{T}\|_{C^2}.$$

In particular, if  $\mathfrak{F}(p, q, \mathbf{v}_p, \mathbf{v}_q) > \frac{4\|\tilde{T}\|_{C^2}}{1-\eta}$ , then

$$\mathfrak{F}(\tilde{T}(p), \tilde{T}(q), D\tilde{T}_p(\mathbf{v}_p), D\tilde{T}_q(\mathbf{v}_q)) \leq \frac{1+\eta}{2} \mathfrak{F}(p, q, \mathbf{v}_p, \mathbf{v}_q).$$

*Proof of Lemma 4.18.* We have

$$\begin{aligned} \mathfrak{F}(\tilde{T}(p), \tilde{T}(q), D\tilde{T}_p(\mathbf{v}_p), D\tilde{T}_q(\mathbf{v}_q)) &= \frac{\angle(D\tilde{T}_p(\mathbf{v}_p), D\tilde{T}_q(\mathbf{v}_q))}{\text{dist}(\tilde{T}(p), \tilde{T}(q))} \\ &\leq \frac{\eta \angle(\mathbf{v}_p, \mathbf{v}_q) + 2\|\tilde{T}\|_{C^2} \text{dist}(p, q)}{\text{dist}(p, q)} \\ &= \eta \mathfrak{F}(p, q, \mathbf{v}_p, \mathbf{v}_q) + 2\|\tilde{T}\|_{C^2}. \end{aligned}$$

If we also have  $\mathfrak{F}(p, q, \mathbf{v}_p, \mathbf{v}_q) > \frac{4\|\tilde{T}\|_{C^2}}{1-\eta}$ , then

$$\begin{aligned} \eta \mathfrak{F}(p, q, \mathbf{v}_p, \mathbf{v}_q) + 2\|\tilde{T}\|_{C^2} &\leq \eta \mathfrak{F}(p, q, \mathbf{v}_p, \mathbf{v}_q) + \frac{1-\eta}{2} \mathfrak{F}(p, q, \mathbf{v}_p, \mathbf{v}_q) \\ &= \frac{1+\eta}{2} \mathfrak{F}(p, q, \mathbf{v}_p, \mathbf{v}_q). \end{aligned}$$

$\square$

Denote  $\Psi = \Phi \circ H$ . From (4.12) we have:

**Lemma 4.19** (Analog of Lemma 3.5 from [12]). *Let  $\tilde{C} > 0$  be given by (4.12). There is  $\nu > 0$  such that the following holds. Suppose that  $a, b \in F_{[0, \nu]}(U_\nu)$ ,  $\mathbf{v}_a \in T_a \mathbb{T}^2$ ,  $\mathbf{v}_b \in T_b \mathbb{T}^2$ . Then the following inequalities hold :*

$$\begin{aligned} \text{dist}(H(a), H(b)) &\leq \tilde{C} \text{dist}(a, b), \\ \text{dist}(a, b) &\leq \tilde{C} \text{dist}(H(a), H(b)), \\ \angle(DH(\mathbf{v}_a), DH(\mathbf{v}_b)) &\leq \tilde{C}(\angle(\mathbf{v}_a, \mathbf{v}_b) + \text{dist}(a, b)), \\ \angle(\mathbf{v}_a, \mathbf{v}_b) &\leq \tilde{C}(\angle(DH(\mathbf{v}_a), DH(\mathbf{v}_b)) + \text{dist}(H(a), H(b))). \end{aligned}$$

Moreover, if  $\Psi(a)$  and  $\Psi(b)$  are defined, then

$$\begin{aligned} \text{dist}(\Psi(a), \Psi(b)) &\leq \tilde{C} \text{dist}(a, b), \\ \text{dist}(a, b) &\leq \tilde{C} \text{dist}(\Psi(a), \Psi(b)), \\ \angle(D\Psi(\mathbf{v}_a), D\Psi(\mathbf{v}_b)) &\leq \tilde{C}(\angle(\mathbf{v}_a, \mathbf{v}_b) + \text{dist}(a, b)), \\ \angle(\mathbf{v}_a, \mathbf{v}_b) &\leq \tilde{C}(\angle(D\Psi(\mathbf{v}_a), D\Psi(\mathbf{v}_b)) + \text{dist}(\Psi(a), \Psi(b))). \end{aligned}$$

From Lemma 4.19 and the definition of  $\mathfrak{F}$  we get the following:

**Lemma 4.20** (Analog of Lemma 3.23 from [12]). *Fix a small  $\nu \geq 0$ . Suppose that  $a, b \in F_{[0, \nu]}(U_\nu)$  are such that  $\Psi(a)$  and  $\Psi(b)$  are defined, and  $\mathbf{v}_a \in T_a \mathbb{T}^2$ ,  $\mathbf{v}_b \in T_b \mathbb{T}^2$ . Then,*

$$\mathfrak{F}(a, b, \mathbf{v}_a, \mathbf{v}_b) \leq \tilde{C}^2(\mathfrak{F}(\Psi(a), \Psi(b), D\Psi(\mathbf{v}_a), D\Psi(\mathbf{v}_b)) + 1)$$

and

$$\mathfrak{F}(\Psi(a), \Psi(b), D\Psi(\mathbf{v}_a), D\Psi(\mathbf{v}_b)) \leq \tilde{C}^2(\mathfrak{F}(a, b, \mathbf{v}_a, \mathbf{v}_b) + 1).$$

Now we are ready to choose a uniform upper bound on the curvatures of all the images of all the curves  $\Gamma_n$  that are outside of small neighborhoods of the singularities. Namely, recall that the constant  $C$  was given by (4.9),  $\tilde{C}$  was defined by (4.12), and define

$$M_{\mathfrak{F}} := \max \left( 2\tilde{C}^2(1 + C), \frac{40\|\tilde{T}\|_{C^2}}{1 - \eta} \right).$$

Choose  $N_0 \in \mathbb{N}$  large enough to make sure that

$$\xi^{-N/4} < \frac{1}{4\tilde{C}^2}.$$

Then for all large  $n$  and for any small piece  $\gamma$  of the curve  $\Gamma_n$ , all the iterates of the curve  $\gamma$  that are in  $U_\nu$  will have curvature not greater than  $M_{\mathfrak{F}}$ . Indeed, after some finite number of initial iterates  $k$ , the curve  $H^{-1}(T^k(\gamma))$  is tangent to the cone field  $K^{\text{cu}}$ , will have curvature bounded by  $M_{\mathfrak{F}}$ , and due to Lemma 4.18 that curvature will remain bounded by  $M_{\mathfrak{F}}$  until we need to change the coordinates by applying the map  $\Psi$ . We will do that only if the image of  $\gamma$  is going to spend more than  $N_0$  iterates in a neighborhood of the singularity. After  $N > N_0$  iterates in a neighborhood of the singularity, due to (4.9) the curvature will be bounded by

$$\xi^{N/4}(\tilde{C}^2(M_{\mathfrak{F}} + 1)) + C \leq \frac{M_{\mathfrak{F}} + 1}{4\tilde{C}^2} + C,$$

and after application of the change of coordinates  $\Psi^{-1}$  we get a curve with curvature bounded by

$$\frac{1}{4}(M_{\mathfrak{F}} + 1) + \tilde{C}^2 C < M_{\mathfrak{F}}.$$

Let us now notice that the partial hyperbolicity of the map  $\tilde{T}$  together with (4.8) implies the following statement (the proof is similar to the proof of [12, Lemma 3.24], so we do not repeat it here):

**Lemma 4.21** (Analog of Lemma 3.24 from [12]). *There is  $R_1 > 0$  such that the following holds for all sufficiently large  $n$ . Suppose that  $\mathbf{v}$  is a non-zero vector tangent to  $\Gamma_n$  at some point  $p \in \Gamma_n$ . Let  $N \in \mathbb{N}$  be such that  $T^N(p)$  belongs to the bounded component of  $S_I \setminus O_{r_1}$ , where  $I = I(p)$ . Then,*

$$\sum_{i=0}^N \|DT^i(\mathbf{v})\| \leq R_1 \|DT^N(\mathbf{v})\|.$$

Lemma 4.21 implies the following statement.

**Lemma 4.22.** *There are constants  $R_1 > 0$  and  $\kappa_1 > 0$  such that for all large  $n$  and any  $N \in \mathbb{N}$ , the following holds. Suppose that  $\gamma \subset T^N(\Gamma_n) \setminus O_{r_1}$  is a connected*

curve of length not greater than  $\kappa_1$ . Let the points  $p, q \in \Gamma_n$  be such that  $T^N(p) \in \gamma$  and  $T^N(q) \in \gamma$ . Then,

$$\sum_{i=0}^N \text{dist}(T^i(p), T^i(q)) < R_1.$$

Finally, since we established a uniform bound on the curvature of all the images of the curves  $\Gamma_n$  outside of small neighborhoods of singularities, we can use Lemma 4.22 together with (4.10) to get the key technical statement that we need to establish the distortion property (i.e., Proposition 4.16):

**Lemma 4.23** (Analog of Lemma 3.19 from [12]). *There are constants  $R > 0$ , and  $\kappa > 0$  such that for any large  $n$  and  $N \in \mathbb{N}$ , the following holds. Suppose that  $\gamma \subset T^N(\Gamma_n) \setminus O_{r_1}$  is a connected curve of length not greater than  $\kappa$ . Let the points  $p, q \in \Gamma_n$  be such that  $T^N(p) \in \gamma$  and  $T^N(q) \in \gamma$ , and  $\mathbf{v}_p$  and  $\mathbf{v}_q$  be unit vectors tangent to  $\gamma$  at points  $p$  and  $q$ . Then*

$$\sum_{i=0}^N (\angle(DT^i(\mathbf{v}_p), DT^i(\mathbf{v}_q)) + \text{dist}(T^i(p), T^i(q))) < R.$$

#### 4.4. Proof of the Distortion Property.

*Proof of Proposition 4.16.* Notice that we need to prove that

$$\left| \log \left( \frac{|\mathfrak{L}_G||\mathfrak{l}_B|}{|\mathfrak{L}_B||\mathfrak{l}_G|} \right) \right|$$

is bounded by some constant independent of the choice of the index  $n$  (as soon as  $n$  is large enough) and the gap in  $\Gamma_n$ . There are points  $p_G \in \mathfrak{l}_G$  and  $p_B \in \mathfrak{l}_B$  such that if  $\mathbf{v}_G$  is a unit vector tangent to the curve  $\mathfrak{l}_G$  at  $p_G$ , and  $\mathbf{v}_B$  is a unit vector tangent to the curve  $\mathfrak{l}_B$  at  $p_B$ , then

$$\begin{aligned} \left| \log \left( \frac{|\mathfrak{L}_G||\mathfrak{l}_B|}{|\mathfrak{L}_B||\mathfrak{l}_G|} \right) \right| &= \left| \log \left( \frac{|T^{n+2}(\mathfrak{l}_G)||\mathfrak{l}_B|}{|T^{n+2}(\mathfrak{l}_B)||\mathfrak{l}_G|} \right) \right| \\ &= \left| \log \left( \frac{|DT^{n+2}(\mathbf{v}_G)|}{|DT^{n+2}(\mathbf{v}_B)|} \right) \right| \\ &= \left| \sum_{i=0}^{n+1} (\log |DT|_{DT^i(\mathbf{v}_G)}(T^i(p_G)) - \log |DT|_{DT^i(\mathbf{v}_B)}(T^i(p_B))) \right| \\ &\leq \sum_{i=0}^{n+1} |\log |DT|_{DT^i(\mathbf{v}_G)}(T^i(p_G)) - \log |DT|_{DT^i(\mathbf{v}_B)}(T^i(p_B))| \\ &\leq \sum_{i=0}^{n+1} ||DT|_{DT^i(\mathbf{v}_G)}(T^i(p_G)) - |DT|_{DT^i(\mathbf{v}_B)}(T^i(p_B))|. \end{aligned}$$

We estimate each of the terms in this sum using the following simple statement applied to the trace map  $T$ :

**Lemma 4.24** (Lemma 3.26 from [12]). *Suppose  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a smooth map,  $a, b \in \mathbb{R}^n$ , and  $\mathbf{v}_a \in T_a \mathbb{R}^n$ ,  $\mathbf{v}_b \in T_b \mathbb{R}^n$  are unit vectors. Then,*

$$||Df|_{\mathbf{v}_a}(a)| - |Df|_{\mathbf{v}_b}(b)|| \leq \|f\|_{C^2}(\angle(\mathbf{v}_a, \mathbf{v}_b) + |a - b|).$$

Now Proposition 4.16 follows from Lemma 4.23.  $\square$

**4.5. Verifying the Assumptions of Propositions 4.2.** Here, we show how to verify the assumptions of Proposition 4.2 in the models under consideration.

*Proof of Theorem 3.4.* Denote the curve of initial conditions by

$$\Gamma(t) = (u_1(t), u_0(t), u_{-1}(t)) \stackrel{\text{def}}{=} (x_1(t^2), x_0(t^2), x_{-1}(t^2)).$$

We then choose  $J_n = [2\pi n + \alpha_n, (2n+1)\pi - \beta_n]$  as above and denote by  $\Gamma_n = \Gamma|_{J_n}$ . We proceed by verifying that conditions (1)–(4) of Proposition 4.2 hold true.

**Property (1).** Write  $v_{-1}(t) = v_0(t) = \cos(t)$  and  $v_1(t) = \cos(2t)$ . Then, put  $u_{j,n} = u_j|_{J_n}$ ,  $v_{j,n} = v_j|_{J_n}$ . First, by [45, Theorem 1.3],

$$\|u_{j,n} - v_{j,n}\|_\infty \lesssim n^{-1},$$

showing that  $u_{j,n} - v_{j,n} \rightarrow 0$  uniformly. In fact, the estimates from [45] hold for complex  $t$  as well. In particular, since all functions in sight are analytic,  $u_{j,n}^{(k)} - v_{j,n}^{(k)} \rightarrow 0$  uniformly for each  $k$  (in particular for  $k = 1, 2$ ) by Cauchy estimates.

**Property (2).** This is immediate from the choice of  $J_n$  and (3.11).

**Property (3).** Note that

$$\frac{d}{dt} \log I(\Gamma_n(t)) = 2 \cot t - 2t^{-1} + \frac{2t \cot \sqrt{t^2 - \lambda}}{\sqrt{t^2 - \lambda}} - \frac{2t}{\sqrt{t^2 - \lambda}}$$

By our choice of  $J_n$ , this is uniformly bounded for large  $n$ .

**Property (4).** This follows immediately from the discussion preceding Proposition 4.2.

Consequently, the result follows from Proposition 4.2 and Lemma 2.3.  $\square$

**Remark 4.25.** In the proof of Theorem 3.4 we made crucial use of the explicit expressions for the traces and the invariant in the case of constant  $f_a, f_b$ . It would of course be of interest to verify the assumptions of Proposition 4.2 for more general choices of  $f_a, f_b$ . For example, one could in this way produce continuous or even smooth potentials. Alas, such an extension will not at all be straightforward. A modified version of property (1) and property (4) can be established in the general case, in fact by the exact same argument, and we expect that property (2) can be established under weaker assumptions as well. However, our proof of property (3) relies on the explicit formula we have for the invariant in the locally constant case, and checking property (3) in a more general case seems to be a challenging technical task.

## 5. THE LOW-ENERGY REGIME

Our goal in this section is to prove Theorem 3.5, which states that in the low-energy regime, the spectrum of the separable 2D Schrödinger operators we consider is a Cantor set. In this proof the existing work on the Fibonacci Hamiltonian will again be helpful, and indeed almost sufficient. As a consequence this section will be less demanding than Section 4.

Our goal is to show that by fixing a suitable compact energy interval  $[0, E_0]$ ,  $E_0 > 0$ , and taking suitable building blocks, we can ensure that

$$(5.1) \quad [0, E_0/2] \cap \Sigma \neq \emptyset$$



and

$$(5.2) \quad [0, E_0] \cap \Sigma = \bigcup_{k=0}^{\infty} A_k, \quad \text{where } A_k \text{ is compact, with } \dim_{\mathbb{B}}^+ A_k < \frac{1}{2}.$$

It is then not hard to verify that

$$(5.3) \quad [0, E_0] \cap (\Sigma + \Sigma) \neq \emptyset$$

and

$$(5.4) \quad \text{Leb}([0, E_0] \cap (\Sigma + \Sigma)) = 0.$$

Since we know that  $\Sigma$  is a Cantor set, it then follows that  $[0, E_0] \cap (\Sigma + \Sigma)$  is compact and nowhere dense. It could happen that  $E_0$  is the edge of a gap of  $\Sigma + \Sigma$  and hence is an isolated point of  $[0, E_0] \cap (\Sigma + \Sigma)$ . However, since  $\Sigma$  itself has no isolated points, this is the only possible isolated point of  $[0, E_0] \cap (\Sigma + \Sigma)$ , and it is easy to avoid this by perturbing  $E_0$ . We will also ensure that  $(-\infty, 0) \cap \Sigma = \emptyset$ , which gives  $(-\infty, 0) \cap (\Sigma + \Sigma) = \emptyset$ . Thus, we will have accomplished our goal in the low-energy region.

Since the primary aim of this paper is to find examples that exhibit certain spectral phenomena, we choose our building blocks in such a way that the proofs of (5.1) and (5.2) are as simple as possible. We expect that the same properties can be shown in greater generality, but only after overcoming significant technical obstacles. To keep the length of the paper in check, we will not pursue this in detail here, but merely point out some of these obstacles in Remark 5.3 below.

One of our building blocks will simply be the zero potential on some finite interval. This allows us to use test functions that are supported in this interval in order to produce low energies in the spectrum, and in particular verify (5.1) if  $E_0$  and the interval length are compatible. Ensuring this is easy: as soon as one of them is fixed, the other one can be determined accordingly.

The choice of the second building block determines the difficulty of proving (5.2). One needs to choose it so that the invariant restricted to  $[0, E_0] \cap \Sigma$  is large. Recall that if the second building block consists of a constant potential, then the invariant can be computed explicitly. Moreover, in this particular case [8] had already proved the desired statement, provided the constant value is sufficiently large. We will recall this derivation below. If, on the other hand, the second building block does not consist of a constant potential, then the invariant can in general not be determined explicitly, and the arguments used in [8] to show that it is large on  $[0, E_0] \cap \Sigma$  (which rely on the explicit formula) cannot be mimicked. In fact, it is not even clear in general how to proceed; see again Remark 5.3 for further comments on the difficulties in establishing (5.2) in the general case.

Let us begin with (5.1).

**Lemma 5.1.** *Suppose  $\ell_a > 0$  and  $f_a \in L^2(0, \ell_a)$  is the zero element. Set*

$$(5.5) \quad E_0 := \frac{24}{\ell_a^2}.$$

*Then (5.1) holds for any choice of  $\ell_b > 0$  and any non-negative  $f_b \in L^2(0, \ell_b)$ .*

*Proof.* Let  $E_0$  be given by (5.5) and suppose that  $\ell_b > 0$  and  $f_b \in L^2(0, \ell_b)$  is non-negative. Fix any  $\omega \in \Omega$ . By (3.7), we have  $\sigma(H_\omega) = \Sigma$ . Thus, our goal is to

show that

$$(5.6) \quad [0, E_0/2] \cap \sigma(H_\omega) \neq \emptyset.$$

We consider the case where  $\omega_0 = \mathbf{a}$ ; the other case is handled similarly (by considering the interval  $[\ell_b, \ell_b + \ell_a)$  instead of the interval  $[0, \ell_a)$ ).

Since  $\omega_0 = \mathbf{a}$ ,  $V_\omega$  vanishes on  $[0, \ell_a)$ . Moreover, the function

$$\varphi(x) = \begin{cases} \frac{2x}{\ell_a} & x \in [0, \frac{\ell_a}{2}) \\ 2 - \frac{2x}{\ell_a} & x \in [\frac{\ell_a}{2}, \ell_a) \\ 0 & \text{otherwise} \end{cases}$$

belongs to the form domain of  $H_\omega$ . We have

$$\langle \varphi, H_\omega \varphi \rangle = \int_0^{\ell_a} |\varphi'(x)|^2 dx = \ell_a \cdot \frac{4}{\ell_a^2} = \frac{4}{\ell_a}.$$

and

$$\|\varphi\|^2 = \int_0^{\ell_a} |\varphi(x)|^2 dx = 2 \int_0^{\frac{\ell_a}{2}} \frac{4x^2}{\ell_a^2} dx = \frac{8}{3\ell_a^2} \left(\frac{\ell_a}{2}\right)^3 = \frac{\ell_a}{3}.$$

Thus,

$$\frac{\langle \varphi, H_\omega \varphi \rangle}{\|\varphi\|^2} = \frac{12}{\ell_a^2},$$

and by the min-max principle (in the form version), it follows that  $\sigma(H_\omega) \cap (-\infty, \frac{12}{\ell_a^2}] \neq \emptyset$ . On the other hand, since  $H_\omega \geq 0$ , we also know that  $\sigma(H_\omega) \cap (-\infty, 0) = \emptyset$ . Together with (5.5) this implies (5.6).  $\square$

Let us now turn our attention to (5.2). As was mentioned above, we do not strive for maximal generality, but rather consider a case where the desired statement can be proved without the need to overcome serious technical difficulties.

**Lemma 5.2.** *Let  $\ell_a = \ell_b = 1$ ,  $f_a = 0 \cdot \chi_{(0,1)}$ , and  $f_b = \lambda \cdot \chi_{(0,1)}$ . Then, (5.2) holds for  $\lambda > 0$  sufficiently large.*

*Proof.* There is  $I_{1/2} \in (0, \infty)$  such that the representation (5.2) of the set  $[0, E_0] \cap \Sigma$  as a countable union of compact sets of small box-counting dimension is available as soon as

$$(5.7) \quad \inf_{E \in \Sigma \cap [0, E_0]} I(E) > I_{1/2}.$$

Indeed, analyticity arguments based on [3] imply that the points of tangency between the curve of initial conditions and the center-stable manifolds of the non-wandering set of the trace map must be isolated, hence there are only finitely many those points in  $\Sigma \cap [0, E_0]$ , if any. If there are no points of tangency, then at every point of  $\Sigma \cap [0, E_0]$  one has

$$\dim_{B, \text{loc}}^+(\Sigma; E) = \dim_{H, \text{loc}}(\Sigma; E).$$

This can be established using the same arguments that were used in the study of the spectrum of the 1D Fibonacci quantum Ising Model in [56]. Concretely, the proof of [56, Theorem 2.1] does not use the specific features of the model, and works for the upper box counting dimension as well, as soon as the curve of initial conditions is transversal to the center-stable manifolds of the non-wandering set of the Fibonacci trace map. The condition (5.7) can in turn be satisfied by choosing

$\lambda$  sufficiently large. This was already discussed in the proof of [8, Corollary 6.7] and we refer the reader to that paper for the details.

In the case when there are points of tangency, denote by  $A_0$  the finite set of energies in  $\Sigma \cap [0, E_0]$  corresponding to points of tangency, and denote by  $A_k$  the set  $(\Sigma \cap [0, E_0]) \setminus U_{\frac{1}{k}}(A_0)$ , where  $U_{\frac{1}{k}}(A_0)$  is the  $\frac{1}{k}$ -neighborhood of  $A_0$ . Then by construction, for each  $k \geq 1$ , at the points parameterized by energies from  $A_k$ , the curve of initial conditions and the center-stable manifolds are transversal, and hence the arguments above imply that  $\dim_{\mathbb{B}}^+ A_k < \frac{1}{2}$  for large  $\lambda$  uniformly in  $k$ . This proves (5.2).  $\square$

With the preparatory work out of the way, we are now in a position to complete the proof of Theorem 3.5.

*Proof of Theorem 3.5.* By Lemmas 5.1 and 5.2, we have (5.1) and (5.2), which in turn imply (5.3) and (5.4). Since  $H_\omega \geq 0$ , it follows that  $(-\infty, E_0] \cap (\Sigma + \Sigma)$  is closed, bounded, nonempty, and has empty interior. Since  $\Sigma$  has no isolated points,  $(-\infty, E_0] \cap (\Sigma + \Sigma)$  also has no isolated points, modulo a small wiggle of  $E_0$ .  $\square$

**Remark 5.3.** Notice that we referred the reader to the proof of [8, Corollary 6.7] for the argument establishing (5.7) in the locally constant case for sufficiently large coupling. Upon inspection of that argument, the reader will notice that any attempt at an extension to a more general setting will face difficulties of various kinds. First of all, the case of locally constant building blocks is such that solutions, and hence transfer matrix traces, and hence the invariant can be computed *explicitly*. The explicit formulae obtained in this way are used in crucial ways in the proof of [8, Corollary 6.7]. Beyond the case of locally constant building blocks, one does in general not have explicit formulae for these quantities. But even if one attempted a generalization of the argument that is based on the qualitative features of these formulae, one would run into difficulties. On a technical level, the problem is related to the study of high barriers in quantum mechanics; see, for example, [28, 33, 53, 54] for previous works on this problem. While the existing papers have investigated the reduction in tunneling upon raising a barrier, as well as resulting localization statements in the presence of a suitable sequence of high barriers, these results appear to be insufficient to produce the precise property we need from a single high barrier in order to derive a statement like (5.7).

## 6. DISCUSSION, QUESTIONS, AND OUTLOOK

We conclude with several questions that we view as natural follow-ups to this work, which nevertheless do not follow from our methods and require additional inspiration.

**Question 1.** Do the results of this paper hold for general  $f_a$  and  $f_b$  satisfying Assumption 3.1? Namely, given  $\ell_\alpha > 0$  and  $f_\alpha \in L^2(0, \ell_\alpha)$ ,  $\alpha \in \{a, b\}$ , is it true that  $\Sigma + \Sigma$  contains a half-line? Writing  $\Sigma_\lambda = \Sigma(\lambda f_a, \lambda f_b)$ , is it true that  $\Sigma_\lambda$  exhibits Cantor structure at low energies for large  $\lambda$ ?

Some of the difficulties one faces when attempting to answer Question 1 were discussed above in Remarks 4.25 and 5.3.

A *Cantorval* is defined to be a nonempty set  $S \subseteq \mathbb{R}$  which is compact, has dense interior, and uncountably many connected components, none of which is isolated.

It is known that  $C + C$  is a Cantorval for “typical” dynamically defined Cantor sets  $C$  [35].

**Question 2.** Is it the case that for some choice of  $f_a$  and  $f_b$ , there is an interval  $J \subseteq \mathbb{R}$  such that  $J \cap (\Sigma + \Sigma)$  is a Cantorval?

Clearly, the study of the structure of the sum of two Cantor sets is crucial to what we do in this paper. It is also important in several other areas and it is therefore a question that has received a lot of attention. There are well-known ways to establish that the sum is a Cantor set (e.g., by bounding the upper box dimension from above, as done in Section 5) or an interval (e.g., by bounding the thickness from below, as done in Sections 2 and 4). While it is known that the sum of two Cantor sets can be a Cantorval [35, 36], the known ways of establishing that (mostly based on [37]) do not seem to be applicable in our setting, where we have far too little control over the relevant properties of the 1D spectra and in which way they depend on the parameters of the model. Thus, an answer to Question 2 appears to be currently completely out of reach.

The Fibonacci sequence can be viewed as a coding of an irrational rotation by  $\theta = \frac{\sqrt{5}-1}{2}$  via:

$$u_n = \begin{cases} \mathbf{a} & n\theta \bmod 1 \in [1-\theta, 1) \\ \mathbf{b} & \text{otherwise.} \end{cases}$$

However, one can naturally replace  $\theta$  by any irrational number. The resulting sequence  $u$  is then called a *Sturmian* sequence.

**Question 3.** What changes if we consider potentials generated by general Sturmian sequences?

In the case when the potentials are generated by a Sturmian sequence, it is still known that  $\Sigma$  is a zero-measure (extended) Cantor set [8]. Moreover, for typical  $\theta$ , there is no hyperbolic dynamical setup analogous to the one used here for  $\theta = \frac{1}{2}(\sqrt{5}-1)$ . Since this setup is absolutely crucial to our approach here, there is a serious obstacle to overcome in generalizing any of our results in this direction for typical  $\theta$ . That being said, in the event that  $\theta$  is a quadratic irrationality, there is a similar (albeit more complicated) formalism [23, 34], so it seems reasonable to expect that many of the techniques of this paper could be applicable in that case as well.

Moreover, it is likely that the results do not in fact hold in full generality for arbitrary frequencies. For instance, in the discrete setting, it is known that the dimension of the spectrum is one for all values of the coupling constant when the frequency is very Liouville [31]. This suggests that for Liouville  $\theta$ , we might not expect to see Cantor structure at the bottom of the spectrum, even for large coupling.

**Question 4.** Can one construct  $V : \mathbb{R}^d \rightarrow \mathbb{R}$  which are almost-periodic and for which the conclusion of Theorem 1.2 holds true?

As mentioned earlier, there is prior work on almost periodic potentials in two dimensions for which the spectrum contains a half line [26, 27]. Is it also possible to prove the Cantor structure of the spectrum at small energies? For non-separable potentials, we are lacking the tools, and for separable potentials, one would need to find 1D almost periodic potentials for which the spectrum is very thin at small

energies; compare the passage from (5.2) to (5.4). In the quasi-periodic case (studied in [27]), no such 1D example is known, and indeed based on the known tools, it is entirely unclear how to produce one. In the limit-periodic case (studied in [26]), there is a dichotomy, which is a function of the existing methods. If the potentials are such that the existence of a half line in the spectrum of the 2D model can be proved, the rate of approximation by periodic potentials is so fast that at small energies, the 1D model has spectrum that is too thick for any known technique to establish the Cantor structure of the sum set. If on the other hand, the rate of approximation is such that the 1D spectrum is sufficiently thin at small energies such that the 2D spectrum has a Cantor structure there, then the spectrum is very thin at all energies and the 2D spectrum is nowhere dense throughout. This mechanism underlies the work [9].

The dichotomy above, showing that one cannot have both desirable properties at the same time, is related to the known phenomenon that for discrete 1D Schrödinger operators with limit-periodic potentials, the spectral type and the qualitative features of the spectrum are the same for all positive values of the coupling constant in all examples in which one is able to prove statements for more than one value of the coupling constant (cf. [7]). This translates to the rough energy-independence of the spectral features discussed above for continuum 1D Schrödinger operators with limit-periodic potentials.

While we focused on the topological structure of the spectrum in the present paper, the following question is naturally of interest as well:

**Question 5.** What can be said about the type of the spectral measures and the density of states measure, both for the specific operators studied in this paper and more generally for multi-dimensional Schrödinger operators with almost periodic or uniformly recurrent potential.

The papers [26, 27] mentioned before also address the spectral type in the high-energy region and prove that it is purely absolutely continuous under their assumptions. Those papers are unable to say anything about the spectral type in the low or medium-energy regime. On the other hand, the operators studied in [9] have purely singular continuous spectrum: the absence of eigenvalues follows from the Gordon lemma, applied to the underlying 1D operators, and the preservation of that fact when forming separable potentials from these 1D pieces; while the absence of absolutely continuous spectrum is immediate from the fact that the spectrum has zero Lebesgue measure. For the operators studied in the present paper, the singular continuous nature of the spectral measure and the density of states measure in the low-energy region follows again from the same arguments, namely Gordon lemma and zero-measure spectrum.<sup>1</sup> We are unable to determine the spectral type of these operators in the large-energy region. Based on analogy with the work [14] on the weakly-coupled discrete case, we would expect that at least the density of states measure is purely absolutely continuous there, but more work will be necessary to prove this. Notice that for similar models, the presence of an absolutely continuous component in the density of states measure in the large-energy region was established in [22]. It is also worth mentioning that the relation between the

---

<sup>1</sup>While [8] did not make the application of the Gordon lemma to the 1D operators studied there explicit, the argument is completely analogous to the one used in the discrete case [16] to exclude eigenvalues.

structure of the spectrum as a set and properties of the density of states measure can be quite non-trivial. For example, in the case of the discrete square Fibonacci Hamiltonian there are regimes with positive measure spectrum but singular density of states measures, see [13]. Similar mechanisms could be present in the continuum case as well. Whether the spectral measures are purely absolutely continuous in the high-energy region for the operators studied here is currently wide open; proving this seems to be well beyond current methods.

Finally we want to mention that, while they do not determine the type of the measures in question, Parnowski and Shterenberg study them in the high-energy regime for quite general multi-dimensional almost-periodic Schrödinger operators in the papers [43, 44], where they obtain complete asymptotic expansions for both the integrated density of states and the spectral function.

## REFERENCES

- [1] J. Avron, B. Simon, Almost periodic Schrödinger operators. I. Limit periodic potentials, *Commun. Math. Phys.* **82** (1981), 101–120.
- [2] M. Baake, U. Grimm, *Aperiodic Order. Vol. 1. A Mathematical Invitation*, Encyclopedia of Mathematics and its Applications **149**, Cambridge University Press, Cambridge, 2013.
- [3] E. Bedford, M. Lyubich, J. Smillie, Polynomial diffeomorphisms of  $\mathbb{C}^2$ . IV. The measure of maximal entropy and laminar currents, *Invent. Math.* **112** (1993), 77–125.
- [4] S. Cantat, Bers and Hénon, Painlevé and Schrödinger, *Duke Math. J.* **149** (2009), 411–460.
- [5] M. Casdagli, Symbolic dynamics for the renormalization map of a quasiperiodic Schrödinger equation, *Comm. Math. Phys.* **107** (1986), 295–318.
- [6] D. Damanik, M. Embree, A. Gorodetski, Spectral properties of Schrödinger operators arising in the study of quasicrystals, in *Mathematics of Aperiodic Order*, 307–370, Progr. Math. **309**, Birkhäuser/Springer, Basel, 2015.
- [7] D. Damanik, J. Fillman, Spectral properties of limit-periodic operators, preprint (arXiv:1802.05794).
- [8] D. Damanik, J. Fillman, A. Gorodetski, Continuum Schrödinger operators associated with aperiodic subshifts, *Ann. Henri Poincaré* **15** (2014), 1123–1144.
- [9] D. Damanik, J. Fillman, A. Gorodetski, Multidimensional almost-periodic Schrödinger operators with Cantor spectrum, *Ann. Henri Poincaré* **20** (2019), 1393–1402.
- [10] D. Damanik, J. Fillman, M. Lukic, Limit-periodic continuum Schrödinger operators with zero-measure Cantor spectrum, *J. Spectral Th.* **7** (2017), 1101–1118.
- [11] D. Damanik, A. Gorodetski, Hyperbolicity of the trace map for the weakly coupled Fibonacci Hamiltonian, *Nonlinearity*, **22** (2009), 123–143.
- [12] D. Damanik, A. Gorodetski, Spectral and quantum dynamical properties of the weakly coupled Fibonacci Hamiltonian, *Commun. Math. Phys.* **305** (2011), 221–277.
- [13] D. Damanik, A. Gorodetski, Spectral transitions for the square Fibonacci Hamiltonian, *Journal of Spectral Theory* **8** (2018), 1487–1507.
- [14] D. Damanik, A. Gorodetski, B. Solomyak, Absolutely continuous convolutions of singular measures and an application to the square Fibonacci Hamiltonian, *Duke Math. J.* **164** (2015), 1603–1640.
- [15] D. Damanik, A. Gorodetski, W. Yessen, The Fibonacci Hamiltonian, *Invent. Math.* **206** (2016), 629–692.
- [16] D. Damanik, D. Lenz, Uniform spectral properties of one-dimensional quasicrystals, I. Absence of eigenvalues, *Commun. Math. Phys.* **207** (1999), 687–696.
- [17] L. H. Eliasson, Floquet solutions for the 1-dimensional quasi-periodic Schrödinger equation, *Commun. Math. Phys.* **146** (1992), 447–482.
- [18] M. Embree, J. Fillman, Spectra of discrete two-dimensional periodic Schrödinger operators with small potentials, *J. Spectral Th.* **9** (2019), 1063–1087.
- [19] P. Exner, O. Turek, Periodic quantum graphs from the Bethe–Sommerfeld perspective, *J. Phys. A: Math. Theor.* **50** (2017), 455201.
- [20] J. Fillman, R. Han, Discrete Bethe–Sommerfeld conjecture for triangular, square, and hexagonal lattices, preprint (arXiv:1806.01988).

- [21] J. Fillman, M. Mei, Spectral properties of continuum Fibonacci Schrödinger operators, *Ann. Henri Poincaré* **19** (2018), no. 1, 237–247.
- [22] J. Fillman, Y. Takahashi, W. Yessen, Mixed spectral regimes for square Fibonacci Hamiltonians, *J. Fract. Geom.* **3** (2016), no. 4, 377–405.
- [23] A. Girand, Dynamical Green functions and discrete Schrödinger operators with potentials generated by primitive invertible substitution, *Nonlinearity* **27** (2014), 527–543.
- [24] R. Han, S. Jitomirskaya, Discrete Bethe–Sommerfeld conjecture, *Commun. Math. Phys.* **361** (2018), 205–216.
- [25] M. Hirsch, C. Pugh, M. Shub, *Invariant Manifolds*, Lecture Notes in Mathematics **583**, Springer-Verlag, Berlin-New York, 1977.
- [26] Y. Karpeshina, Y.-R. Lee, Spectral properties of a limit-periodic Schrödinger operator in dimension two, *J. Anal. Math.* **120** (2013), 1–84.
- [27] Y. Karpeshina, R. Shterenberg, Extended states for the Schrödinger operator with quasi-periodic potential in dimension two, *Mem. Amer. Math. Soc.* **258** (2019), no. 1239.
- [28] W. Kirsch, S. Molchanov, L. Pastur, One-dimensional Schrödinger operators with high potential barriers, *Operator Calculus and Spectral Theory (Lambrecht, 1991)*, 163–170, Oper. Theory Adv. Appl. **57**, Birkhäuser, Basel, 1992.
- [29] S. Klassert, D. Lenz, P. Stollmann, Delone measures of finite local complexity and applications to spectral theory of one-dimensional continuum models of quasicrystals, *Discrete Contin. Dyn. Syst.* **29** (2011), 1553–1571.
- [30] H. Krüger, Periodic and limit-periodic discrete Schrödinger operators, *preprint* (arXiv:1108.1584).
- [31] Q. Liu, Y. Qu, Z. Wen, The fractal dimensions of the spectrum of Sturm Hamiltonian, *Adv. Math.* **257** (2014), 285–336.
- [32] C. Matheus, The Lagrange and Markov Spectra from the Dynamical Point of View. In: Ferenczi S., Kulaga-Przymus J., Lemanczyk M. (eds) *Ergodic Theory and Dynamical Systems in their Interactions with Arithmetics and Combinatorics. Lecture Notes in Mathematics*, vol. 2213 (2018), Springer, Cham.
- [33] I. McGillivray, P. Stollmann, G. Stolz, Absence of absolutely continuous spectra for multi-dimensional Schrödinger operators with high barriers, *Bull. London Math. Soc.* **27** (1995), 162–168.
- [34] M. Mei, Spectra of discrete Schrödinger operators with primitive invertible substitution potentials, *J. Math. Phys.* **55** (2014), no. 8, 082701.
- [35] C. Moreira, E. Morales, J. Rivera-Letelier, On the topology of arithmetic sums of regular Cantor sets, *Nonlinearity* **13** (2000), 2077–2087.
- [36] C. Moreira, E. Morales, Sums of Cantor sets whose sum of dimensions is close to 1, *Nonlinearity* **16** (2003), 1641–1647.
- [37] C. Moreira, J.-C. Yoccoz, Stable intersections of regular Cantor sets with large Hausdorff dimensions, *Ann. of Math.* **154** (2001), 45–96.
- [38] J. Moser, An example of a Schrödinger equation with almost periodic potential and nowhere dense spectrum, *Comment. Math. Helv.* **56** (1981), 198–224.
- [39] S. Newhouse, Nondensity of axiom A(a) on  $S^2$ , *Global Analysis (Proc. Sympos. Pure Math., Vol. XIV, Berkeley, Calif., 1968)*, 191–202, Amer. Math. Soc., Providence, 1970.
- [40] S. Newhouse, The abundance of wild hyperbolic sets and nonsmooth stable sets for diffeomorphisms, *Inst. Hautes Études Sci. Publ. Math.* **50** (1979), 101–151.
- [41] J. Palis, F. Takens, *Hyperbolicity and Sensitive Chaotic Dynamics at Homoclinic Bifurcations*, Cambridge University Press, 1993.
- [42] L. Parnowski, Bethe–Sommerfeld conjecture, *Ann. Henri Poincaré* **9** (2008), 457–508.
- [43] L. Parnowski, R. Shterenberg, Complete asymptotic expansion of the integrated density of states of multidimensional almost-periodic Schrödinger operators, *Ann. of Math. (2)* **176** (2012), 1039–1096.
- [44] L. Parnowski, R. Shterenberg, Complete asymptotic expansion of the spectral function of multidimensional almost-periodic Schrödinger operators, *Duke Math. J.* **165** (2016), 509–561.
- [45] J. Pöschel, E. Trubowitz, *Inverse Spectral Theory*, Academic Press, Inc., Boston, MA, 1987.
- [46] C. Pugh, M. Shub, A. Wilkinson, Hölder foliations, *Duke Math. J.* **86** (1997), 517–546.
- [47] M. Reed, B. Simon, *Methods of Modern Mathematical Physics. I. Functional Analysis*, 2nd edition, Academic Press, New York, 1980.



- [48] M. Reed, B. Simon, *Methods of Modern Mathematical Physics. IV. Analysis of Operators*, Academic Press, New York-London, 1978.
- [49] B. Simon, On the genericity of nonvanishing instability intervals in Hill's equation, *Ann. Inst. H. Poincaré Sect. A (N.S.)* **24** (1976), 91–93.
- [50] M. Skriganov, Proof of the Bethe–Sommerfeld conjecture in dimension two, *Soviet Math. Dokl.* **20** (1979), 89–90.
- [51] M. Skriganov, Geometric and arithmetic methods in the spectral theory of multidimensional periodic operators, *Proc. Steklov Math. Inst.* **171** (1984), 3–122.
- [52] M. Skriganov, The spectrum band structure of the three-dimensional Schrödinger operator with periodic potential, *Inv. Math.* **80** (1985), 107–121.
- [53] P. Stollmann, G. Stolz, Singular spectrum for multidimensional Schrödinger operators with potential barriers, *J. Operator Theory* **32** (1994), 91–109.
- [54] G. Stolz, Localization for Schrödinger operators with effective barriers, *J. Funct. Anal.* **146** (1997), 416–429.
- [55] O. Veliev, Spectrum of multidimensional periodic operators, *Teor. Funktsii Funktsional. Anal. i Prilozhen* **49** (1988), 17–34 (in Russian).
- [56] W. Yessen, On the spectrum of 1D quantum Ising quasicrystal, *Ann. Henri Poincaré* **15** (2014), 419–467.

DEPARTMENT OF MATHEMATICS, RICE UNIVERSITY, HOUSTON, TX 77005, USA  
*E-mail address:* `damanik@rice.edu`

DEPARTMENT OF MATHEMATICS, TEXAS STATE UNIVERSITY, SAN MARCOS, TX 78666  
*E-mail address:* `fillman@txstate.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, IRVINE, CA 92697, USA,  
 AND NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS, RUSSIAN FEDERATION  
*E-mail address:* `asgor@uci.edu`